PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/00 | A2

(11) International Publication Number: **WO 00/09676**

(43) International Publication Date: 24 February 2000 (24.02.00)

(21) International Application Number: PCT/US99/18525

(22) International Filing Date: 13 August 1999 (13.08.99)

(30) Priority Data:
60/096,630 14 August 1998 (14.08.98) US

(71) Applicant: CALIFORNIA INSTITUTE OF TECHNOLOGY [US/US]; 1200 East California Boulevard, Pasadena, CA 91125 (US).

(72) Inventors: ANDERSON, David, J.; 2891 Mount Curve Avenue, Pasadena, CA 91001 (US). LO, Li-Ching; 525 West Las Tunas Drive, Arcadia, CA 91006 (US).

(74) Agents: TRECARTIN, Richard, F. et al.; Flehr Hohbach Test Albritton & Herbert LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHODS OF FORMING NEURONS

(57) Abstract

The invention relates to novel methods of inducing non-neuronal cells to differentiate into neurons and to methods of inducing non-neuronal cells to express a neuronal subtype-specific marker.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHODS OF FORMING NEURONS

This is a continuing application of United States Application No. 60/096,630, filed 08/14/98, pending.

FIELD OF INVENTION

This invention relates to the expression of transcription factors in non-neuronal cells to induce their differentiation into neuronal cells, and more particularly, to the use the members of the neurogenin family to induce neurogenesis and the use of Phox2a or b to produce neurons which express tyrosine hydroxylase.

BACKGROUND OF THE INVENTION

Differentiation of uncommitted neuronal precursors cells into neurons is regulated by the coordinated expression in a cascade fashion of a variety of transcription factors. Transcription factors are proteins that recognize and bind to specific DNA sequences located in or around chromosomal genes, and thereby regulate the expression of those genes by increasing or decreasing their rate of transcription. There are dozens of different "families" of transcription factors, members of which share a common specificity for a given DNA recognition sequence, for example, homeodomain proteins, zinc finger proteins, and basic helix-loop-helix (bHLH) proteins. Within these families there are scores of different proteins which have related but distinct structures, and different patterns of expression.

In regards to nerve cell differentiation, transcription factors have been identified that induce precursor cells to commit to neuronal differentiation or induce committed cells to express properties shared by a specific type or subtype of neuron. For example,

mammalian homologs of the *Drosophila* proneural genes, called MASH1 and NEUROGENINS (NGNs)-1, -2, and -3 (Johnson *et al.* Nature 1990. 346:858-861; Ma *et al.* Cell. 1996. 87:43-52; Sommer *et al.* Cell. Neurosci. 1996. 8:221-224) are expressed in neuronal precursors (Ma *et al.* J. Neurosci. 1997. 17:3644-3652) and are required for commitment to a neuronal fate (Ma *et al.* Neuron. 1998 20:469-482). Studies in *Xenopus* suggest that the NGNs regulate a core program of neurogenesis, that is shared by many different classes of neurons in the central nervous system (CNS) and peripheral nervous system (PNS) (Ma *et al.* Cell. 1996. 87:43-52; Bellefroid *et al.* Cell. 1996. 87:1191-1202). (MASH1 may play a similar, although not identical, role for those neurogenic lineages that do not utilize NGNs (Guillemot *et al.* Cell. 1993. 75:463-476; Lo *et al.* Curr. Biol. 1997. 7:440-450.) In addition, paired-like homeodomain protein, Phox2a (Valarché *et al.* Development. 1993. 119:881-896) (and a close relative, Phox2b), is expressed by a more restricted subset of neurons in the CNS and PNS than express the NGNs or MASH1. In particular, expression of Phox2 proteins correlates with expression of a noradrenergic neurotransmitter phenotype as well as with expression of c-RET, a signal transducing receptor for Glial cell line-Derived Neurotrophic Factor (GDNF) (Tiveron *et al.* J. Neurosci. 1996. 16:7649-7660).

It has also been reported that the NeuroD family of transcription factors are involved in neurogenesis. In particular, it has been reported that NeuroD7 and NeuroD2 expression in *Xenopus* embryos induced neurogenesis in ectodermal cells (McCormick *et al.*, Mol. Cell Biol. 1996. 16(10):5792-5800). However, the McCormick study reports that the ectopic neurons induced by NeuroD7 and NeuroD2 were confined to a subpopulation of ectodermal cells, as shown, by spotty NCAM-positive staining pattern. McCormick further reports that the apparent restricted activity of the NeuroD proteins to a subset of cells derived from the ectoderm suggests that other factors may regulate their activity, such as, the notch pathway that mediates lateral inhibition during *Drosophila* neurogenesis.

Recent studies have explored the extent to which the differentiation of neuronal precursor cells, neural crest stem cells (NCSC), to particular neuronal phenotypes can

be controlled by forced expression of these transcription factors. These studies have demonstrated that forced expression of MASH1 using retroviral vectors can induce some, but not all, NCSCs to differentiate into neurons (Lo *et al.*, Development. 1998. 125:609-620). These neurons express some markers common to all neurons, such as neurofilament, and others that are specific to autonomic neurons in the PNS. The latter include the aforementioned transcription factor Phox2a and the receptor c-RET, an autonomic subtype marker. (Lo *et al.* Development. 1998. 125:609-620). Forced expression of Phox2a in NCSCs, like MASH1, led to induction of c-RET; however unlike MASH1, Phox2a was unable to promote the core program of neurogenesis (Lo *et al. supra*). These data support the idea that the differentiation of neural stem cells to particular neuronal subtypes is controlled by a combination of transcription factors, some of which promote a core program of neurogenesis and others of which promote expression of neuronal subtype-specific properties.

Therefore, it is an object herein to provide the compounds necessary to promote a core program of neurogenesis as well as those necessary to promote expression of neuronal subtypes. It is also an object to provide methods for inducing non-neuronal cells to differentiate into neurons.

An additional object herein to provide methods for controlling neural stem cell differentiation in order to generate neural cells of a particular phenotype in quantities suitable for transplantation. In contrast to most other fully differentiated cells, neurons lose their capacity to regenerate and, therefore, congenital defects, diseases or trauma to central and peripheral nervous systems, such as, blindness, deafness, neurodegenerative disorders, Parkinson's Disease, Huntington's Disease, and Multiple Sclerosis, and damage or trauma associated with encephalitis or injury are difficult to correct. Furthermore, tumors in neural tissues can also be very difficult to treat because of the toxic side effects that conventional chemotherapeutic drugs may have on nervous tissues. Surgical removal of tumors may also result in neuronal damage. Accordingly, it is an object herein to achieve controlled differentiation of neural stem cells of the CNS into dopaminergic neurons, for use in transplantation therapies of Parkinson's

Disease, of GABAergic striatal interneurons for therapy of Huntington's Disease, or of oligodendrocytes for therapy of Multiple Sclerosis.

A further object herein to use such neuronal differentiating agents and information provided herein for construction of test cell lines, animal models, assays for identifying candidate agents which modulate neurogenesis, assays for identifying therapeutic agents, gene therapy, and differentiation of tumor cells.

SUMMARY

In accordance with the objects outlined above, the present invention provides methods for inducing a non-neuronal cell to differentiate into a neuronal cell through the recombinant expression of a transcription factor that induces a core program of neurogenesis

In another aspect, the present invention provides methods for inducing the expression of a neuronal subtype-specific marker in a non-neuronal cell.

In a further aspect, the invention provides expression vectors comprising transcriptional and translational control sequences operably linked to a nucleic acid encoding a member of the neurogenin family of transcription factors or a Phox2a or Phox2b transcription factor, and host cells containing the expression vector(s).

In an additional aspect, the invention provides cells having an induced neuronal phenotype comprising an expression vector comprising transcriptional and translational control sequences operably linked to a nucleic acid encoding a member of the neurogenin family. The invention also provides cells that have been induced to express a neuronal subtype-specific marker comprising an expression vector comprising transcriptional and translational control sequences operably linked to a nucleic acid encoding a Phox2a or Phox2b protein.

In a further aspect, the invention provides for identifying agents that modulate the

induction of a core program of neurogenesis and/or a neuronal subtype specific marker.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A shows the rat neurogenin-1 (NGN1) nucleic acid sequence (SEQ ID NO:1).

Figure 1B shows the rat NGN1 amino acid sequence (SEQ ID NO:2).

Figure 1C shows the mouse NGN1 nucleic acid sequence (SEQ ID NO:3).

Figure 1D shows the mouse NGN1 amino acid sequence (SEQ ID NO:4).

Figure 1E shows the xenopus X-ngr-1a cDNA sequence (SEQ ID NO:5).

Figure 1F shows the xenopus X-ngr-1a amino acid sequence (SEQ ID NO:6).

Figure 1G shows the xenopus X-ngr-1b cDNA sequence (SEQ ID NO:7).

Figure 1H shows the xenopus X-ngr-1b amino acid sequence (SEQ ID NO:8).

Figure 1I shows the mouse NGN2 nucleic acid (SEQ ID NO:9) and amino acid (SEQ ID NO:10) sequences.

Figure 1J shows the mouse NGN3 nucleic acid (SEQ ID NO:11) and amino acid sequences (SEQ ID NO:12).

Figure 2A shows the mouse Phox2a nucleic acid sequence (SEQ ID NO:13). (Valarche, *et al.* 1993. Development. 1993, 119:881-886).

Figure 2B shows the mouse Phox2a deduced amino acid sequence (SEQ ID NO:14). The homeodomain (HD) is underlined. (Valarche, *et al.* 1993. Development. 1993, 119:881-886).

Figure 2C shows the most Phox2b nucleic acid sequence (SEQ ID NO:15). (Pattyn et al. (1997) Development 124(20):4065-4075).

Figure 2D shows the most Phox2b amino acid sequence (SEQ ID NO:16). (Pattyn et al. (1997) Development 124(20):4065-4075).

Figures 3A-C show the induction of neuronal differentiation in NCSCs by the bHLH transcription factor, MASH1, expressed from a MASH1-IRES-GFP encoding retrovirus. Figure 3A shows NCSCs infected with MASH1-IRES-GFP retrovirus have a process-bearing neuronal morphology. Figure 3B shows NCSCs infected with MASH1-IRES-GFP retrovirus express NF160. Figure 3C shows NCSCs infected with MASH1-IRES-GFP retrovirus are identified by GFP fluorescence.

Figures 4A-C show induction of neuronal differentiation in NCSCs by the bHLH transcription factor NGN1 expressed from an NGN1-IRES-GFP retrovirus. Figure 4A shows all NCSCs infected with an NGN1-IRES-GFP retrovirus have a neuronal morphology. Figure 4B shows all NCSCs infected with NGN1-IRES-GFP retrovirus stain positively with anti-NF160 antibody. Figure 4C shows all NCSCs infected with NGN1-IRES-GFP retrovirus are identified by GFP fluorescence.

Figures 5A-C show induction of neuronal differentiation of NCSCs grown at high density by NGN1-IRES-GFP retrovirus. Figure 5A shows the morphology of NCSCs grown at high density and infected with NGN1-IRES-GFP retrovirus. Figure 5B shows that NCSCs grown at high density and infected with NGN1-IRES-GFP retrovirus stain positively with an anti-NeuN antibody. Figure 5C shows that NCSCs grown at high density and infected with NGN1-IRES-GFP retrovirus can be identified by GFP fluorescence.

Figures 6A-D show induction of a neuronal marker in cultured chick embryo fibroblasts infected with RCAS replication-competent retrovirus expressing NGN1 containing the myc epitope tag. Figure 6A shows chick embryo fibroblasts infected with the RCAS

replication-competent retrovirus expressing NGN1 containing the myc epitope tag stain positively with anti-myc tag antibody. Figure 6B shows chick embryo fibroblasts infected with the RCAS replication-competent retrovirus expressing NGN1 containing the myc epitope tag stain positively with antibody 3A10 which recognizes a neurofilament-associated protein, NAPA-73. Figure 6C shows chick embryo fibroblasts infected with the RCAS replication-competent retrovirus expressing NGN1 containing the myc epitope tag stain positively with antibody 270RMO which recognizes NF160. Figure 6D shows chick embryo fibroblasts infected with the RCAS replication-competent retrovirus expressing NGN1 containing the myc epitope tag stain positively with antibody TuJ1 which recognizes beta-III tubulin. Insets of Figures 6B, 6C, 6D shows higher magnification of their respective figures.

Figure 7 shows the effect of added factors on TH induction by Phox2a or GFP retroviral infected NCSCs. Abbreviations: no add=no added factor; GDNF=glial cell line-derived Neurotrophic Factor; Dex=dexamethasone; F+G+D=forskolin + GDNF + Dex.

Figure 8 shows the effect of different factors on the percentage of TH+ NCSCs in all Phox2a retrovirus infected myc+ clones. Abbreviations: no add=no added factor; GDNF=glial cell line-derived Neurotrophic Factor; Dex=dexamethasone; F+G+D=forskolin + GDNF + Dex.

Figures 9A-D shows the effect of induced TH expression by NCSCs infected with a retrovirus expressing myc epitope tagged Phox2a protein and cultured with added factors, forskolin, FGF (fibroblast growth factor), and dexamethasone (F+G+D). Figure 9A shows that Phox2a-myc tagged retrovirus infected NCSCs have a non-neuronal morphology. Figure 9B shows that Phox2a-myc tagged retrovirus infected NCSCs positively with anti-myc-tag antibody. Figure 9C shows induced TH expression by fluorescent staining in NCSCs infected with Phox2a-myc tagged retrovirus. Figure 9D shows a double exposure of Figs 9B and 9C to demonstrate that

many Phox2a-myc expressings NDSCs co-express TH.

Figures 10A-F show a comparison of NCSCs infected with either Phox2a-myc tagged retrovirus or GFP-myc tagged retrovirus. Figure 10A shows that NCSCs infected with Phox2a-myc tagged retrovirus and treated with forskolin have a non-neuronal morphology. Figure 10B shows that NCSCs infected with GFP-myc tagged retrovirus and treated with forskolin have a non-neuronal morphology. Figure 10C shows that NCSCs infected with Phox2a-myc tagged retrovirus and treated with forskolin stain positively with anti-myc-tag antibody. Figure 10D shows that NCSCs infected with GFP-myc tagged retrovirus and treated with forskolin stain positively with anti-myc-tag antibody. Figure 10E shows that NCSCs infected with Phox2a-myc tagged retrovirus and treated with forskolin stain positively with anti-TH antibody. Figure 10F shows that NCSCs infected with GFP-myc-tagged retrovirus and treated with forskolin do not stain with anti-TH antibody.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel methods for inducing non-neuronal cells to differentiate into neurons. The present invention also provides novel methods for inducing a non-neuronal cell to express a neuronal-subtype specific marker.

In a preferred embodiment, a vector encoding a transcription factor is introduced into a non-neuronal host cell. The transcription factor is operably linked to a promoter and transcription termination regulatory sequences active in the host cell. Expression of the encoded transcription factor induces the non-neuronal host cell to differentiate into a neuron. Alternatively, expression of the transcription factor induces the non-neuronal host cell to express a neuronal subtype-specific marker.

Accordingly, the invention provides transcription factors that can be expressed in a non-neuronal host cell and which induces the host cell to differentiate into a neuronal cell or express a neuronal sub-type specific marker. By "induce" herein is meant to cause a host cell to express at least one endogenous gene. Preferably, inducing herein

refers to producing a neuronal phenotype in a cell not showing such a phenotype prior to expression of a transcription factor. By "transcription factor" as used herein is meant a protein that regulates the transcription and expression of a gene or genes in a host cell. In one embodiment, the transcription factor induces a core program of neurogenesis. In another embodiment, the transcription factor induces a host cell to express a neuronal subtype-specific marker. By a "core program of neurogenesis" herein is meant the induced expression of a marker or markers common to all neurons. Examples include a process bearing neuronal morphology, or the expression of neurofilament protein, neuron-specific nucleoprotein, neuron-specific beta-tubulin, or NF160. It is believed that the temporal aspect of the expression of neurogenins and the phenotype of NGN knockouts contributes to the characterization of neurogenins as being the primary initiator of neural differentiation and the induction of a cascade of genes that induce neural differentiation. By a "neuronal subtype-specific marker or property" herein is meant a marker or property associated with only a subset of neurons, such as, tyrosine hydroxylase (TH) expression.

In a preferred embodiment, a host cell is induced to express a core program of neurogenesis and a neuronal sub-type specific marker by the expression of a combination of the appropriate transcription factors.

In one embodiment a transcription factor of the present invention includes members of the neurogenin family. By "neurogenin" herein is meant a transcription factor, such as neurogenin-1 (NGN1), neurogenin-2 (NGN2), or neurogenin-3 (NGN3) that is expressed in non-neuronal cells and induces a core program of neurogenesis.

In another embodiment a transcription factor of the present invention includes, for example, Phox2. By "Phox2" herein is meant a transcription factor, such as Phox2a, that induces the expression of properties associated with a specific subtype of neuron, such as neurons that synthesize the catecholamine family of neurotransmitters which include dopamine, noradrenaline (norepinephrine), and adrenaline (epinephrine). For example, Phox2a induces the expression of TH which is the rate-limiting enzyme in the

synthesis of catecholamines. In the peripheral nervous system (PNS), TH is expressed by sympathetic autonomic neurons, and in the central nervous system (CNS) by dopaminergic neurons of the Substantia Nigra, noradrenergic neurons of the Locus Coeruleus, and several other groups of neurons. Other transcription factors which may be used in accordance with the invention include POU homeodomain proteins (e.g. Brn-3.0/3a); paired homeodomain proteins (e.g. DRG-11); LIM homeodomain proteins (e.g. Isl-1, Lhx-3); Nkx-family homeodomain proteins (Dlx-1, -2, Nkx2.1, 2.2, 2.5 etc.); zinc finger protein (GATA-2, -3); bHLH proteins (eHAND, dHAND); orphan nuclear receptors (e.g., Nurr-1); homeodomain proteins such as MNR2 or HB9.

By "non-neuronal cell" herein is meant any cell that is not a neuron. Therefore, a non-neuronal cell is any cell that does not function as a conducting cell of the peripheral or central nervous system. Accordingly, a non-neuronal cell includes uncommitted neuron progenitors or precursor cells, for example, neural crest stems cells (NCSC) or neural stem cells (NSC). Non-neuronal cells also includes glia (astrocytes, Schwann cells, oligodenocrocytes) cells that are not of a neuronal origin or lineage and include, for example, fibroblasts. Other cells that may be used in accordance with the invention include, for example, embryonic stem cells (ES cells); neural stem cells derived from ES cells; mesenchymal stem cells; satellite cells; sustentacular cells; endocrine cells; epidermal stem cells; muscle stem cells; neuroepithelial precursor (NEP) cells; neuroblastoma cells and other cells types may be used.

By "neural crest stem cell" herein is meant a cell derived from the neural crest which is characterized by having the properties (1) of self-renewal and (2) asymmetrical division; that is, one cell divides to produce two different daughter cells with one being self (renewal) and the other being a cell having a more restricted developmental potential, as compared to the parental neural crest stem cell. The foregoing, however, is not to be construed to mean that each cell division of a neural crest stem cell gives rise to an asymmetrical division. It is possible that a division of a neural crest stem cell can result only in self-renewal, in the production of more developmentally restricted progeny only, or in the production of a self-renewed stem

cell and a cell having restricted developmental potential. The neural crest gives rise to the peripheral nervous system (PNS).

By the term "neural stem cell" refers to a multipotent cell having properties similar to that of a neural crest stem cell but which is not necessarily derived from the neural crest. Rather, as described hereinafter, such multipotent neural stem cells can be derived from various other tissues including neural epithelial tissue from the brain and/or spinal cord of the adult or embryonic central nervous system or neural epithelial tissue which may be present in tissues comprising the peripheral nervous system. In addition, such multipotent neural stem cells may be derived from other tissues such as lung, bone and the like. In a preferred embodiment, multipotent neural stem cells are derived from the PNS, such as from the neural crest, and not from the CNS. It is to be understood that such cells are not limited to multipotent cells but may comprise a pluripotent cell capable of regeneration and differentiation to different types of neurons and glia, *e.g.*, PNS and CNS neurons and glia or progenitors thereof. In this regard, it should be noted that the neural crest stem cells described herein are at least multipotent in that they are capable of self-regeneration and differentiation to some but not all types of neurons and glia *in vitro*. Thus, a neural crest stem cell is a multipotent neural stem cell derived from a specific tissue, *i.e.*, the embryonic neural tube or the sciatic nerve (Morrison et al. 1999).

The transcription factors of the present invention may be identified in several ways, including, by substantial nucleic acid or amino acid sequence similarity or identity to the sequences shown in Figures 1A-L and Figure 2A-B. Sequence similarity or identity can be based upon the overall nucleic acid or amino acid sequence. The transcription factors of the present invention have been found in vertebrates including zebrafish (*Danio)*, mice (*Mus*), rats (*Rattus*), birds (*Gallus*) and amphibians (*Xenopus*), and it is therefore expected to be found in a number of organisms, such as zebrafish and primates.

As used herein, a protein is a "neurogenin protein" if the overall similarity of the

protein sequence to the amino acid sequence of the neurogenin depicted herein is preferably greater than about 85%, more preferably greater than about 90% and most preferably greater than about 95%. In some embodiments the similarity will be as high as about 98-99%.

In addition, a neurogenin protein preferably also has a neurogenin basic-helix-loop-helix (bHLH) domain, which comprises a DNA-binding and dimerization domain (Johnson *et al.* Nature. 1990. 356:858-861).

As used herein, a protein is a "Phox2 protein" if the overall similarity of the protein sequence to the amino acid sequences of the Phox2a depicted herein is preferably greater than about 85%, more preferably greater than about 90% and most preferably greater than 95%. In some embodiments the similarity will be as high as about 98-99%.

In addition, a Phox2a protein preferably also has a homeodomain (HD). (Valarche, *et al.* 1993. Development. 1993, 119:881-886)

The transcription factors of the present invention may be shorter or longer than the amino acid sequences shown in the Figures 1A-L and Figure 2A-B. Thus, in a preferred embodiment, included within the definition of transcription factors of the present invention are portions or fragments of the sequences depicted herein. Generally fragments have up to about 100-150 residues, with about 15 to about 50 residues being preferred, and about 50 to about 100 residues being more preferred and about 100-150 being most preferred. Fragments of the transcription factor proteins are considered transcription factor proteins if one or more of the following characteristics exist: a) they share at least one antigenic epitope; b) have at least the indicated similarity; c) and preferably have a biological activity associated with the full length sequence. "Biological activity" includes the ability to induce a core program of neurogenesis or induce a neuronal subtype-specific phenotype.

Similarity is determined using standard techniques known in the art, including, but not limited to, the algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the algorithm of Needleman & Wunsch. J. Mol. Biol. 1970. 48:443, by the search for similarity method of Pearson & Lipman. 1988. PNAS USA 85:2444, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Drive, Madison, WI), or the Best Fit sequence program described by Devereux *et al.* Nucl. Acid Res. 1984. 12:387-395.

In a preferred embodiment, percent identity or similarity is calculated by FastDB based upon the following parameters: mismatch penalty of 1.0; gap penalty of 1.0; gap size penalty of 0.33, joining penalty of 30.0. ("Current Methods in Comparison and Analysis", Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp. 127-149, 1998. Alan R. Liss, Inc.)

Another example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng and Doolittle. J. Mol. Evol. 1987. 35:351-360; the method is similar to that described by Higgins and Sharp. 1989. CABIOS 5:151-153. Useful PILEUP parameters including a default gap weight of 3.00, a default gap length weight of 0.10, and weighted end gaps.

An additional example of a useful algorithm is the BLAST algorithm, described in Altschul *et al.* J. Mol. Biol. 1990. 215:403-410 and Karlin *et al.*, PNAS USA 1993. 90:5873-5787. A particularly useful BLAST program is the WU-BLAST-2 program which was obtained from Altschul *et al.*, Methods in Enzymology. 1996. 266: 460-480; [http://blast.wustl/edu/blast/ README.html]. WU-BLAST-2 uses several search parameters, most of which are set to the default values. The adjustable parameters are

set with the following values: overlap span =1, overlap fraction = 0.125, word threshold (T) = 11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched; however, the values may be adjusted to increase sensitivity.

An additional useful algorithm is gapped BLAST as reported by Altschul *et al.* Nucleic Acids Res. 25:3389-3402. Gapped BLAST uses BLOSUM-62 substitution scores; threshold *T* parameter set to 9; the two-hit method to trigger ungapped extensions; charges gap lengths of *k* a cost of 10+*k*; X_u set to 16, and X_g set to 40 for database search stage and to 67 for the output stage of the algorithms. Gapped alignments are triggered by a score corresponding to ~22 bits.

In an alternative embodiment, percent amino acid sequence identity is determined. In percent identity calculations relative weight is not assigned to various manifestations of sequence variation, such as, insertions, deletions, substitutions, *etc.* Only identities are scored positively (+1) and all forms of sequence variation given a value of "0", which obviates the need for a weighted scale or parameters as described above for sequence similarity calculations. Therefore, percent identity represents a highly rigorous method of comparing sequences.

Percent sequence identity can be calculated, for example, by dividing the number of matching identical residues by the total number of residues of the "shorter" sequence in the aligned region and multiplying by 100. The "longer" sequence is the one having the most actual residues in the aligned region.

By "neurogenin nucleic acid" or "Phox2a nucleic acid" is meant, respectively, a nucleic acid encoding a neurogenin protein or Phox2a protein, as defined herein. Nucleic acids encoding the transcription factors of the present invention can be identified by a number of methods as known in the art.

In one embodiment, the neurogenin or Phox2a nucleic acids are identified by sequence similarity as outlined below. In the case of nucleic acids encoding the transcriptioin factors of the present invention the overall similarity of the nucleic acid sequence is commensurate with the amino acid similarity of the encoded transcription factor but takes into account the degeneracy in the genetic code and codon bias of different organisms. As will be appreciated by those in the art, due to the degeneracy of the genetic code, large numbers of nucleic acids may be made, all of which encode the transcription factors of the present invention. Thus, having identified a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids, by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the encoded protein. Accordingly, the nucleic acid sequence similarity may be either lower or higher than that of the protein sequence. Thus the similarity of the nucleic acid sequences encoding the transcription factors of the present invention as compared to the nucleic acid sequences of Figures 1A-L and 2A-B are preferably greater than 60%, more preferably greater than about 70%, particularly greater than about 75% and most preferably greater than 80%. In some embodiments the similarity will be as high as about 90 to 95 or 98%.

Nucleic acid similarity can be determined using, for example, BLASTN (Altschul *et al.* 1990. J. Mol. Biol. 147:195-197). BLASTN uses a simple scoring system in which matches count +5 and mismatches -4. To achieve computational efficiency, the default parameters have been incorporated directly into the source code.

In another embodiment, the nucleic acid similarity is determined through hybridization studies. Thus, for example, nucleic acids which hybridize under high stringency to the nucleic acid sequences shown in the Figures (SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15) and their complements are considered neurogenin or Phox2 genes. High stringency conditions are known in the art; *see* for example Maniatis *et al.*, Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, *et al.*, Hames and Higgins, eds. *Nucleic Acid Hybridization, A Practical Approach*, IL press, Washington, D.C., 1985; Berger and Kimmel eds. *Methods in*

Enzymology, Vol. 52, *Guide to Molecular Cloning Techniques*, Academic press Inc., New York, N.Y., 1987; and Bothwell, Yancopoulos and Alt, eds, *Methods for Cloning and Analysis of Eukaryotic Gene*, Jones and Bartlett Publishers, Boston, Mass. 1990, which are hereby expressly incorporated by reference in their entirety.

The choice of hybridization conditions will be evident to one skilled in the art and will generally be guided by the purpose of the hybridization, the type of hybridization (DNA-DNA, DNA-RNA, RNA-RNA, oligonucleotide-DNA *etc.*), and the level of desired relatedness between the sequences. Methods for hybridization are well established in the literature. For example, one or ordinary skill in the art realizes that the stability of nucleic acid duplexes will decrease with an increased number and proximity of mismatched bases; thus, the stringency of hybridization may be used to maximize or minimize the stability of such duplexes. Hybridiziation stringency can be altered by, for example, adjusting the temperature of hybridization solution; adjusting the percentage of helix-destabilizing agents, such as, formamide, in the hybridization solution; and adjusting the temperature and salt concentration of the wash solutions. In general, the stringency of hybridization is adjusted during the post-hybridization washes by varying the salt concentration and/or the temperature. Stringency of hybridization may be increased, for example, by: i) increasing the percentage of formamide in the hybridization solution; ii) increasing the temperature of the wash solution; or iii) decreasing the ionic strength of the wash solution. High stringency conditions may involve high temperature hybridization (*e.g.* 65°C-68°C in aqueous solution containing 4-6X SSC, or 42°C in 50% formamide) combined with high temperature (e.g., 5°C-25°C below the T_m) and a low salt concentration (e.g., 0.1X SSC) washes. Reduced stringency conditions may involve lower hybridization temperatures (*e.g.*, 35°C-42°C in 20-50% formamide) with intermediate temperature (*e.g.*, 40°C-60°C) washes in a higher salt concentration (e.g., 2-6X SSC). Moderate stringency condtions, which may involve hybridization at a temperature between 50°C-55°C and washes in 0.1X SSC, 0.1% SDS at between 50°C and 55°C, may be used (*see* Maniatis and Ausubel, *supra*). In a preferred embodiment, nucleic acids which hybridize to the nucleic acids herein have the biological activity as described herein.

The transcription factor encoding nucleic acids of the present invention are preferably recombinant. As used herein, "nucleic acid" may refer to either DNA or RNA, or molecules which contain both deoxy- and ribonucleotides. The nucleic acids include genomic DNA, cDNA and oligonucleotides including sense and anti-sense nucleic acids. Such nucleic acids may also contain modifications in the ribose-phosphate backbone to increase stability and half life of such molecules in physiological environments. The recombinant nucleic acids of the present invention may be double stranded, single stranded, or contain portions of both double stranded or single stranded sequence.

By the term "recombinant nucleic acid" herein is meant nucleic acid, originally formed *in vitro*, in general, by the manipulation of nucleic acid by endonucleases, polymerases, ligases, and/or recombinases to produce a form not normally found in nature. Alternatively, a recombinant nucleic acid may be chemically synthesized according to organic synthesis methods. Thus, a recombinant nucleic acid of the present invention encodes a transcription factor that induces neurogenesis in non-neuronal cells and is distinguished from the corresponding transcription factor-encoding nucleic acid molecule as it exists in natural or unmodified cells.

Accordingly, a recombinant nucleic acid of the present invention can be in a linear or circular form. Following introduction of a recombinant nucleic acid into a host cell the nucleic acid can reside in a host cell as an extrachromosomal element or can be incorporated into the genome of a host cell. A host cell can have one or multiple copies of the recombinant nucleic acid extrachromosomally or inserted into the host cell genome. In an alternative embodiment, a host cell may have both extrachromosomal and inserted forms.

It is understood that once a recombinant nucleic acid is made and introduced into a host cell or organism, it will replicate non-recombinantly, i.e. using the *in vivo* cellular machinery of the host cell rather than *in vitro* manipulations; however, such nucleic acids, once produced recombinantly, although subsequently replicated non-

recombinantly, are still considered recombinant for the purposes of the invention.

A "recombinant protein" is a protein made using recombinant techniques. A recombinant protein is distinguished from naturally occurring protein by at least one or more characteristics. For example, a recombinant protein is expressed from a recombinant nucleic acid, such as an expression vector, as described below. As such, the definition of a recombinant protein includes a transcription factor protein of the present invention produced from a recombinant nucleic acid either *in vitro*, *in vivo*, or *in situ*. The recombinant protein or transcription factor can be from one organism but is expressed in a different organism or host cell. The level or degree of expression of the recombinant transcription factor may be higher or lower than is normally seen. To regulate the level of expression the use of an inducible promoter may be used. In an alternative embodiment, the transcription factor may be in a form not normally found in nature, as in the addition of an epitope tag or amino acid substitutions, insertions and deletions, as discussed below.

In a preferred embodiment, expression of the recombinant protein is at least sufficient to induce the differentiation of the host cell into a neuron. In an alternative preferred embodiment, the expression of the recombinant protein is at least sufficient to induce the expression of a neuronal subtype-specific marker.

In a preferred embodiment, a recombinant nucleic acid is an expression vector. By "expression vector" herein is meant a nucleic acid that encodes and directs the synthesis of a transcription factor of the present invention. Expression of the transcription factor is effected by operably linking the sequence encoding the transcription factor to control sequences.

The term "control sequences" refers to sequences necessary for the expression of an operably linked coding sequence *in vitro*, *in vivo*, or *in situ*. The control sequences that are suitable for non-neuronal cell expression, in general, include but are not limited to, promoter sequences, ribosomal binding sites, transcriptional start and stop sequences,

translational start and stop sequences, and enhancer or activator sequences. In a preferred embodiment, the regulatory sequences include a promoter and transcriptional start and stop sequences.

Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous. However, enhancers do not have to be contiguous, as described below. Linking the sequence encoding the transcription factor to control sequences is generally accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice. Alternatively, linking can be accomplished by employing mutagenesis techniques, PCR, recombination, organic synthesis methods, or a combination of these methods, as known in the art.

A promoter is any nucleic acid sequence for all cell types including eukaryotic and prokaryotic cells as known in the art capable of binding a RNA polymerase and initiating the downstream (3') transcription of a coding sequence for a transcription factor protein into mRNA. In a preferred embodiment, a promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, using a located 25-30 base pairs upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase to begin RNA synthesis at the correct site. A eukaryotic promoter from a cell or virus may also contain an upstream promoter element (enhancer element), typically located within 100 to 200 base pairs upstream of the TATA box, as described below. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation. Of particular use as eukaryotic promoters are the promoters from viral genes, since viral genes are often highly expressed and have a broad host range. Examples include the bovine papilloma virus promoter, SV40 early

promoter, avian sarcoma virus LTR promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter, hepatitis-B virus promoter, fowlpox virus (UK 2,211,504 published 5 July 1989), herpes simplex virus promoter, and the CMV promoter. Examples of eukaryotic promoters from mammalian cells include, the actin promoter or an immunoglobulin promoter, and heat-shock promoters, provided such promoters are compatible with the host cell systems. Preferably, the promoter chosen is functional in the non-neuronal cell of choice so as to control expression of the neurogenenin or Phox2a genes.

In a preferred embodiment, the promoter activity is inducible or can be modulated, such as an ecdysone-inducible promoter-enhancer combination, an estrogen-induced promoter-enhancer combination, a tetracycline-inducible promoter-enhancer, a CMV promoter-enhancer, an insulin gene promoter, or other cell-type specific, developmental stage-specific, hormone-inducible, factor-inducible, or drug inducible, promoter. When a hormone- or factor-inducible promoter is used, the cell must have the required hormone or factor receptor present, either naturally or as a consequence of expression of a co-transfected expression vector encoding such receptor. Accordingly, the host cell must be responsive to the hormone or factor that regulates the corresponding promoter's activity.

In contrast to the naturally occurring promoters, described above, alternatively the promoters can be hybrids of two or more promoters. Hybrid or compound promoters, which contain elements of more than one promoter, are known in the art, and are useful in the present invention. Examples of hybrid promoters include the Tetracycline Responsive Element/minimal immediate early promoter of cytomegalovirus. Such promoters can be designed to be active in the presence or absence of tetracycline. (*see* Clontech 98/99 Catalog, Palo Alto, CA, which is expressly incorporated by reference in its entirety).

Transcription of a nucleic acid encoding the transcription factor may be increased by inserting an enhancer sequence into the vector. Enhancers are *cis*-acting elements of

DNA, usually about from 10 to 300 bp, that act on a promoter to increase its transcription. Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, alpha-fetoprotein, and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers. The enhancer may be spliced into the vector at a position 5' or 3' to the transcription factor encoding sequence, but is preferably located at a site 5' from the promoter.

Expression vectors of the present invention will also contain sequences necessary for the termination of transcription and for stabilizing the mRNA. Such sequences are commonly available from the 5' and, occasionally 3', untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding the transcription factor. Typically, transcription termination and polyadenylation sequences are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-translational cleavage and polyadenylation. Examples of eukaryotic transcription terminator and polyadenlytion signals include those derived form SV40, herpes simplex virus, retroviral 3'-LTR, the beta-globin gene, and the bovine growth hormone gene.

The expression vector may comprise additional elements. For example, the expression vector may have two replication systems, thus allowing it to be maintained in two organisms, for example, in eukaryotic cells for expression and induction of neurogenesis and in a procaryotic host for cloning and amplification. Furthermore, for integrating expression vectors, the expression vector contains at least one sequence homologous to the host cell genome, and preferably two homologous sequences which flank the expression construct (the nucleic acid encoding the transcription factor operably linked to control sequences). The integrating vector may be directed to a

specific locus in the host cell by selecting the appropriate homologous sequence for inclusion in the vector. Constructs for integrating vectors are well known in the art.

Preferably, expression vectors will typically contain selection gene(s), also termed a selectable marker, for selection in eukaryotic and for prokaryotic cells, as needed. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, neomycin, hygromycin, puromycin, bleomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, *e.g.*, the gene encoding D-alanine racemase for *Bacilli.*

A further example of suitable selectable markers for mammalian cells are those that enable the identification of cells competent to take up the transcription factor(s)-encoding nucleic acid, such as DHFR or thymidine kinase. An appropriate host cell when wild-type DHFR is employed is deficient in DHFR activity, prepared and propagated as described by Urlaub *et al., Proc. Natl. Acad. Sci. USA, 77*:4216 (1980).

Still other vectors suitable for adaptation to the synthesis of transcription factors in recombinant vertebrate cell culture are described in Gething *et al., Nature, 293*:620-625 (1981); Mantei *et al., Nature, 281*:40-46 (1979); EP 117,060; EP 117,058; Clontech 98/99 (Palo Alto, CA), Promega 1998 (Madison, WI); and Life Technologies 97/98 (Gaithersburg, MD) catalogs.

Accordingly, the expression vector, for example, may be in the form of a plasmid or viral particle. Examples of plasmid expression vectors include pTargeT™, pSI, pCI (Promega, Madison, WI); pSV●Sport (Life Technologies, Gaithersburg, MD); pTRE (Clontech, Palo Alto, CA). Viral expression systems include retroviruses (pBABE), adenoviruses, herpesviruses (McLean *et al. JID 170*(5):1100-1109 (1994), and togaviruses (Sindbis and Semliki Forest viruses). In a preferred embodiment, the viral vector is deficient in one or more essential genes and is replication-incompetent in target host cells. Preferered vectors are retroviral expression vectors, for example,

pBABE and others, which are preferrably inducible.

In a further embodiment, the transcription factors of the present invention may also be made as a fusion protein, using techniques well known in the art. Thus, for example, the expressed transcription factor protein may be fused to a carrier protein or polypeptide, for example, in order to modulate the level of expression, biological activity, or monitor the expressed factor. For example, the transcription factor can be fused to the Antp homeodomain (A. Prochiantz. (1998) Nature Biotechnol. 16:819-820; Derossi et al. (1998) Trends Cell Biol. 8:84-87).

To monitor expression, the transcription factor can be fused to a polypeptide that functions as an epitope tag. The epitope tag is generally placed at the amino- or carboxyl- terminus of the transcription factor. The presence of such epitope-tagged forms of the transcription factor can be detected using an antibody against the tag polypeptide. Also, provision of the epitope tag enables the transcription factor to be readily purified by affinity purification using an anti-tag antibody or another type of affinity matrix that binds to the epitope tag. Various tag polypeptides and their respective antibodies are well known in the art. Examples include poly-histidine (poly-his) or poly-histidine-glycine (poly-his-gly) tags; the flu HA tag polypeptide and its antibody 12CA5 (Field *et al., Mol. Cell. Biol., 8*:2159-2165 (1988)]; the c-myc tag and the 8F9, 3C7, 6E10, G4, B7 and 9E10 antibodies thereto [Evan *et al., Molecular and Cellular Biology*, 5:3610-3616 (1985)); and the Herpes Simplex virus glycoprotein D (gD) tag and its antibody [Paborsky *et al., Protein Engineering, 3*(6):547-553 (1990)]. Other tag polypeptides include the Flag-peptide [Hopp *et al., BioTechnology, 6*:1204-1210 (1988)]; the KT3 epitope peptide [Martin *et al., Science, 255*:192-194 (1992)]; an a-tubulin epitope peptide [Skinner *et al., J. Biol. Chem., 266*:15163-15166 (1991)]; and the T7 gene 10 protein peptide tag [Lutz-Freyermuth *et al., Proc. Natl. Acad. Sci. USA, 87*:6393-6397 (1990)].

In a further embodiment, the transcription factors of the present invention may also be made as amino acid sequence variants. These variants fall into one or more of three

classes: substitutional, insertional or deletional variants. In one embodiment, these variants are prepared by site specific mutagenesis of nucleotides in the DNA encoding the transcription factor protein, using cassette, or PCR mutagenesis or other techniques well known in the art to produce DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture as outlined below to identity the variant with the desired properties. However, variant transcription factor protein fragments having up to about 100-150 residues may be prepared by *in vitro* synthesis using established organic synthesis methods techniques.

Amino acid sequence variants are characterized by the nature of the variation, a feature that sets them apart from naturally occurring allelic or interspecies variation of the transcription factor amino acid sequence. The variants typically exhibit the same qualitative biological activity as the naturally occurring analogue, although variants can also be selected which have modified characteristics as will be more fully outlined below.

In certain embodiments, when the site or region for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed transcription factor variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence are well known, for example, M13 primer mutagenesis and PCR mutagenesis. Screening of the mutants is done using assays of transcription factor protein activities. Amino acid substitutions are typically of single residues; insertions usually will be on the order of from about 1 to 20 amino acids, although considerably larger insertions may be tolerated. Deletions range from about 1 to about 20 residues, although in some cases deletions may be much larger.

Substitutions, deletions, insertions or any combination thereof may be used to arrive at a final derivative. Generally these changes are done on a few amino acids to minimize

the alteration of the molecule. However, larger changes may be tolerated in certain circumstances. When small alterations in the characteristics of the transcription factor protein are desired, substitutions are generally made in accordance with the following chart:

Original Residue	Exemplary Substitutions
Ala	Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those shown in Chart I. For example, substitutions may be made which more significantly affect: the structure of the polypeptide backbone in the area of the alteration, for example the alpha-helical or beta-sheet structure; the charge or hydrophobicity of the molecule at the target site; or the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in the polypeptide's properties are those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g.

lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g. glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g. phenylalanine, is substituted for (or by) one not having a side chain, e.g. glycine.

In one embodiment, variants exhibit the same qualitative biological activity and will elicit the same immune response as the naturally-occurring analogue. In an alternative embodiment, variants are selected to modify the characteristics of the transcription factor proteins as needed, such as, the biological activity and/or immunogenic properties, as described below.

In an alternative embodiment, a library of variants are generated by an entirely, non-specific, random mutagenesis method. These techniques are known in the art and do not require the selection of a specific cite or region to be altered. For example, DNA shuffling as described by Stemmer. *Nature 370*:389-391 (1994) and Stemmer. *PNAS USA 91*:10747-10751 (1994)) can be used to produce variants which are cloned, expressed, and screened for a desired property. For example, the intracellular activity of the transcription factor can be increased or decreased. In addition, the number and types of genes that are regulated by the transcription factor can also be broadened or narrowed, as needed to induce expression within a host cell, as described below.

Also included with the definition of transcription factor variants are transcription factor proteins from other organisms, which are cloned and expressed as outlined below. Thus, probe or degenerate polymerase chain reaction (PCR) primer sequences may be used to find other related transcription factor proteins from humans or other organisms. As will be appreciated by those in the art, particularly useful probe and/or PCR primer sequences include the less conserved areas and preferably, the unique areas of the nucleic acid sequence encoding the transcription factor proteins of the present invention. As is generally known in the art, preferred PCR primers are from about 15 to about 35 nucleotides in length, with from about 20 to about 30 being preferred, and may contain inosine as needed. The conditions for the PCR reaction are well known in the art. It is therefore also understood that provided along with the sequences in the

sequences listed herein are portions of those sequences, wherein unique portions of 15 nucleotides or more are particularly preferred. The skilled artisan can routinely synthesize or cut a nucleotide sequence to the desired length.

The methods of introducing the expression vectors into target host cells, are well known in the art, and will vary with the host cell and the type of expression vector that is used. The target host cell can be in tissue culture or, alternatively, can be in an organism. For DNA or RNA vectors, techniques include the use of dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the expression vector into cell or nuclei. For the case of recombinant virus particle vectors, entry into a host cell is mediated via attachment of the virus particle to the host cell followed by penetration of the host cell membrane and introduction of the viral nucleocapsid into the host cell. The mechanism of entry will vary according to the type of virus vector being used but generally will follow the mechanisms of entry of wild-type virus.

Transformed host cells of the present invention find a variety of *in vitro* uses, for example: i) as convenient sources of neuronal and other growth factors, ii) in transient and continuous culture for screening drugs or compounds that are either antagonists or protagonists of neural differentiation as it relates to normal differentiation and development, neural repair, and tumor development, iii) as sources of recombinantly expressed neurogenins and/or Phox2a proteins for use as an antigen in preparing monoclonal and polyclonal antibodies useful in diagnostic assays, iv) in transient and continuous cultures for screening for compounds capable of increasing or decreasing the activity of neurogenin and/or Phox2a, vi) for use in transplantation, as described below and, vii) *in vivo* delivery of the genes into adult neural stem cells to induce neurogenesis *in vivo*.

For expression in host cells, specific conditions may vary with the cell type being used and the desired neuron or neuronal subtype-specific marker to be produced. The

transcription factors of the present invention are produced by culturing a host cell transformed with an expression vector containing nucleic acid encoding a neurogenin or Phox2a protein, under the appropriate conditions cause expression of the encoded factor. The conditions appropriate for the specific transcription factor(s) expression will vary with the choice of the expression vector and the host cell, and will be easily ascertained by one skilled in the art through routine experimentation. For example, the use of constitutive promoters in the expression vector will require optimizing the growth and proliferation of the host cell, while the use of an inducible promoter requires the appropriate growth conditions for induction. In addition, in some embodiments, the timing of the harvest is important.

In a preferred embodiment, expression of the transcription factors of the present invention in a non-neuronal host cell induces the cell to differentiate into a neuron. In an another embodiment, expression of the transcription factors of the present invention induce the cell to express and neuronal subtype-specific marker. In yet another embodiment, expression of the transcription factors of the present invention in a non-neuronal host cell induces the cell to differentiate into a neuron and to express a neuronal subtype-specific marker. Appropriate host cells for the induction of a neuronal phenotype/expression include, for example, neural stem cells, neural crest stem cells, and cells of a non-neuronal origin or lineage, such as, fibroblasts or epithelial cells or as described *supra*. Especially preferred cells are embryonic stem cells.

In a preferred embodiment, expression of neurogenin in non-neuronal cells and uncommitted neuronal precursor cells, such as, neural stem cells or neural crest stem cells or fibroblasts induces a core program of neurogenesis associated with the commitment of a cell to differentiate into a neuron cell. The core program of neurogenesis include a number of markers common to all neurons. Examples of these markers include adoption of a neuronal morphology, or expression of neurofilament, neuron-specific nucleoprotein, neuron-specific beta-tubulin, NF160, NeuN, SCG10, neuron-specific enolase, PGP9.5, hi-PSA NCAM, synapsin I.

In another preferred embodiment, expression of Phox2a in non-neuronal cells and uncommited neuronal precursor cells induces the expression of properties associated with specific neuronal subtype, for example, neurons that synthesize catecholamine neurotransmitters which include dopamine, noradrenaline, and adrenaline. Phox2a preferably induces the expression of tyrosine hydrolase (TH) which is the rate-limiting enzyme in the synthesis of catecholamines. In the PNS, TH is expressed by sympathetic autonomic neurons, and, noradrenergic neurons of the Locus Coeruleus, and several other groups of neurons. It is therefore desirable to be able to control the differentiation of neurons that express TH, from neural stem cells.

In yet another preferred embodiment, expression of neurogenin and Phox2a in non-neuronal cells induces the expression of both a core program of neurogenesis and properties associated with a neurons that synthesize catecholamines.

The cells of the present invention also find a variety of *in vivo* uses, for example, for transplantation at sites of neuronal disfunction. For example, cells are transformed *in vitro* and are transplanted into an organism. In a preferred embodiment, the transplanted host cells replace or enhance functions of neurons that communicate via electrical or chemical synapses. Examples of neurons that communicate via chemical synapses include, for example, peptidergic, serotonergic, noradrenergic, cholinerigc, glutamatergic, GABAergic, dopaminergic, and noradrenergic neurons.

In a preferred embodiment the transplantation is autologous but alternatively can be heterologous or a xenographic transplant. For other than autologous transplantation, immune suppressors or modifiers are preferably employed, as known in the art, to prevent destruction of the transplanted cells or tissue by a host verses graft response.

The transformed cells are transplanted in a quantity to be therapeutically effective. A therapeutically effective quantity or dosage refers to a dosage adequate to ameliorate symptoms or signs of the disease without producing unacceptable toxicity to the patient. In general, an effective quantity of transplanted cell is that which provides

either subjective relief of symptoms or an objectively identifiable improvement as noted by the clinician or other qualified observer.

The dosage or quantity of transplanted cells used in accordance with this invention varies depending on the cell and the condition being treated. The age, weight, and clinical condition of the recipient patient, and the experience and judgment of the clinician, practitioner, or veterinarian administering the therapy are among the factors affecting the selected dosage. Other factors include the patient's medical history, the severity of the disease process, and the potency of the particular transplanted cells.

The host cells can be transplanted to either the central or peripheral nervous system. The central nervous system (CNS) includes, for example, the cortex, hippocampus, septum, striatum, the cerebrum, cerebellum, pons, medulla oblongata, neural tissues of the pituitary gland, the spinal cord etc. The peripheral nervous system (PNS) includes all neural tissue that is not a component of the CNS.

In another aspect of the invention, the expression vectors are introduced into cells *in vivo*. Accordingly, induction of a core program of neurogenesis or a neuronal subtype-specific marker occurs *in vivo* in cells containing an expression vector of the present invention. The compositions for administration will preferably comprise a solution of the expression vector dissolved or suspended in a pharmaceutically acceptable carrier. The type of pharmaceutically acceptable carrier will be directed, in part, by the type of expression vector that is employed, for example, a nucleic acid vector or a viral vector. A variety of carriers can be used, e.g, buffered saline containing suitable emulsifiers, and the like. Methods of producing liposomes and complexing or encapsulating compounds therein are well known to those of skill in the art *(see, e.g.*, Debs and Zhu (1993) WO 93/24640; Mannino and Gould-Fogerite (1988) *BioTechniques* 6(7): 682-691; Rose U.S. Pat No. 5,279,833; Brigham (1991) WO 91/06309; and Felgner *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84: 7413-7414).

In a preferred embodiment, it is desirable to package, complex, or otherwise combine

the expression vector with a delivery vehicle that preferably increases cellular uptake and/or half-life. A wide variety of suitable vehicles are well known to those of skill in the art. Thus, for example, the expression vector can be complexed with, or encapsulated within, a charged lipid to form a net neutral composition. This will reduce clearance by the reticuloendothelial system and enhance cellular uptake.

In another embodiment, the expression vector can be encapsulated within or complexed with microparticles which can be recognized and phagocytosed by a target cell thereby facilitating entry of the expression vector into the cell. Other methods of facilitating entry include the use fusion proteins, protein complexes, and masking charged groups with reversible chemical modification or counterions. Viral vectors, such as, adenovirus, retroviruses (e.g. lentivirus), herpesviruses (e.g. herpes simplex virus), togaviruses (e.g., Sindbis virus), also can be used.

For certain of the therapeutic uses of the subject expression vectors, particularly peripheral uses such as for induction of neurogenesis in the peripheral nervous system, direct (*e.g.*, topical or injected) administration of the expression vector will be appropriate, according to the type of expression vector that is employed. Accordingly, the subject expression vector, alone or in combination with a delivery vehicle may be conveniently formulated for administration with a biologically acceptable medium, such as water, buffered saline, polyol (for example, glycerol, propylene glycol, liquid polyethylene glycol and the like) or suitable mixtures thereof. In preferred embodiments, the expression vector is dispersed in lipid formulations, such as micelles, which closely resemble the lipid composition of natural cell membranes to which the expression vector is to be delivered.

As indicated above, the expression vectors are preferably combined with a pharmaceutically acceptable carrier for *in vivo* administration. Pharmaceutically acceptable carriers (excipients) can contain a physiologically acceptable compound that acts, for example, to solubilize the composition, and/or to stabilize the composition, and/or to increase or decrease the absorption of the agent. Physiologically acceptable

compounds can include, for example, carbohydrates, such as glucose, sucrose, or dextrans, antioxidants, such as ascorbic acid or glutathione, chelating agents, low and/or high molecular weight proteins, compositions that reduce the clearance or hydrolysis of the expression vectors, or excipients or other stabilizers and/or buffers. Other physiologically acceptable compounds include wetting agents, emulsifying agents, dispersing agents or preservatives which are particularly useful for preventing the growth or action of microorganisms.

The expression vector pharmacological compositions are preferably sterile and generally free of undesirable matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents; toxicity adjusting agents and the like, for example, sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate and the like.

The concentration of expression vector in these formulations can vary widely, and will be selected primarily based on fluid volumes, viscosities, body weight and the like in accordance with the particular mode of administration selected and the patient's needs.

Where the expression vector is used in a therapeutic context, (*e.g.*, in the treatment of a condition characterized by neuronal disfunction or deficiency), a therapeutically effective quantity of expression vector is employed in treatment. A therapeutically effective quantity or dosage refers to a dosage adequate to ameliorate symptoms or signs of the disease without producing unacceptable toxicity to the patient. In general, an effective amount of the compound is that which provides either subjective relief of symptoms or an objectively identifiable improvement as noted by the clinician or other qualified observer.

The dosage of expression vector compositions used in accordance with this invention varies depending on the compound and the condition being treated. The age, weight,

and clinical condition of the recipient patient, and the experience and judgment of the clinician, practitioner, or veterinarian administering the therapy are among the factors affecting the selected dosage. Other factors include the route of administration, the patient's medical history, the severity of the disease process, and the potency of the particular compound.

Representative patient populations that may benefit from transplantation include: patients with hearing or vision loss due to optical or auditory nerve damage, patients with central or peripheral nerve damage and loss of motor or sensory neural activity, patients with brain or spinal cord damage, patients with neurodegenerative disease or disorders. The damage may be the result of trauma and can be induced by injury, accident, stroke (infarction, ischemia, hypoxia) or medical treatment, for example, surgery, or may represent a congenital birth defect, for example, paralysis, blindness, or deafness. The damage may also be the result of an autoimmune disease or the sequelae of an infectious disease, for example, meningitis, encephalitis, human immunodeficiency virus, and prions.

Representative neurodegenerative diseases and disorders that are treated or ameliorated by the transformed host cells of the present invention include, for example, Alzheimer's Disease, Amyotrophic Lateral Sclerosis (ALS), Huntington's Disease (HD), Multiple Sclerosis (MS), Parkinson's Disease (PS), and Epilepsy.

The transformed host cells of the present invention also find use in the identification of compounds or candidate bioactive agents, such as, proteins including polypeptides and oligopeptides, lipids, carbohydrates, nucleic acids, including oligonucleic acid and antisense nucleic acids, small organic molecules, inorganic molecules, steroids, *etc.* that modulate the activity of the transcription factors of the present invention. Thus, in one embodiment, this invention provides methods of identifying transcription factor modulators that specifically block or enhance transcription factor activity.

The methods involve screening the "candidate compound's" ability to modulate

induction of a core program of neurogenesis and/or a neuronal subtype-specific property in host cells transformed with an expression vector of the present invention. The host cell can be a cell in culture, in an organism, or, alternatively transgenic animals may be used.

Screens may be designed to first find candidate agents that can bind to transcription factor proteins, and then these agents may be used in assays that evaluate the ability of the candidate agent to modulate transcription factor protein activity. Thus, as will be appreciated by those in the art, there are a number of different assays which may be run; binding assays and biological activity assays.

Thus, in a preferred embodiment, the methods comprise combining transcription factor protein and a candidate bioactive agent, and determining the binding of the candidate agent to the transcription factor protein. Preferred embodiments utilize the transcription factor proteins as described herein but , although other transcription factor proteins may also be used, including rodents (mice, rats, hamsters, guinea pigs, etc.), farm animals (cows, sheep, pigs, horses, etc.) and primates (humans). These latter embodiments may be preferred in the development of animal models of human disease. In some embodiments, as outlined herein, variant or derivative transcription factor proteins may be used, as outlined above.

The term "candidate bioactive agent" or "exogeneous compound" as used herein describes any molecule, e.g., protein, polypeptide, oligopeptide, lipids, carbohydrates, nucleic acids, oligonucleic acid, including antisense nucleic acids, small organic molecules, inorganic molecules, steroids, *etc.*, with the capability of directly or indirectly altering the biological activity of transcription factor protein. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration or below the level of detection.

Modulation of transcription factor biological activity is indicated at the first detectable level. A change in activity, which can be an increase or decrease, is preferably a change of at least 20% to 50%, more preferably by at least 50% to 75%, more preferably at least 75% to 100%, and more preferably 150% to 200%, and most preferably is a change of at least 2 to 10 fold compared to a control.

Nucleic acids which encode transcription factor protein or its modified or variant forms can also be used to generate either transgenic animals or "knock out" animals which, in turn, are useful in the development and screening of therapeutically useful compounds or reagents. A transgenic animal (e.g., a mouse or rat) is an animal having cells that contain a transgene, which transgene was introduced into the animal or an ancestor of the animal at a prenatal, e.g., an embryonic stage. A transgene is a DNA which is integrated into the genome of a cell from which a transgenic animal develops. In one embodiment, cDNA encoding a transcription factor protein can be used to clone genomic DNA encoding an transcription factor protein in accordance with established techniques and the genomic sequences used to generate transgenic animals that contain cells which express the desired DNA from a transgene. Methods for generating transgenic animals, particularly animals such as mice or rats, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009. Typically, particular cells would be targeted for the transcription factor protein transgene incorporation with tissue-specific enhancers. Transgenic animals that include a copy of a transgene encoding transcription factor protein introduced into the germ line of the animal at an embryonic stage can be used to examine the effect of increased expression of the desired nucleic acid. Such animals can be used as tester animals for reagents thought to confer protection from, for example, pathological conditions associated with its overexpression. In accordance with this facet of the invention, an animal is treated with the reagent and a reduced incidence of the pathological condition, compared to untreated animals bearing the transgene, would indicate a potential therapeutic intervention for the pathological condition.

Alternatively, a transcription factor protein "knock out" animal which has at least one

defective, deleted, or altered allele encoding a transcription factor protein as a result of homologous recombination between the endogenous gene encoding a transcription factor protein and altered genomic DNA encoding a transcription factor protein introduced into an embryonic cell of the animal. For example, cDNA encoding an transcription factor protein can be used to clone genomic DNA encoding a transcription factor protein in accordance with established techniques. A portion of the genomic DNA encoding a transcription factor protein can be deleted or replaced with another gene, such as a gene encoding a selectable marker which can be used to monitor integration. Typically, several kilobases of unaltered flanking DNA (both at the 5' and 3' ends) are included in the vector [see e.g., Thomas and Capecchi, Cell, 51:503 (1987) for a description of homologous recombination vectors]. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced DNA has homologously recombined with the endogenous DNA are selected [see e.g., Li et al., Cell, 69:915 (1992)]. The selected cells are then injected into a blastocyst of an animal (e.g., a mouse or rat) to form aggregation chimeras [see e.g., Bradley, in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E. J. Robertson, ed. (IL, Oxford, 1987), pp. 113-152]. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term to create a "knock out" animal. Progeny harboring the homologously recombined DNA in their germ cells can be identified by standard techniques and used to breed animals in which all cells of the animal contain the homologously recombined DNA or in which both alleles are defective, deleted, or altered. Knockout animals can be characterized for instance, for their life-expectency and cause of death, their ability to defend against certain pathological conditions and for their development of pathological conditions due to absence of the transcription factor protein polypeptide. For example, knockouts in NGN-1, -2, MASH1, Phox 2a, Phox 2b, and combinations thereof are made, for example, in mice. It is understood that cell based knock-out or "knock-in" systems can also be made and utilized in accordance with the present disclosure.

It is understood that the models described herein can be varied. For example, "knock-in" models can be formed, or the models can be cell-based rather than animal models.

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes. All cited/referenced patents, patent applications, publications and references cited therein are expressly incorporated by reference in their entirety.

EXAMPLES

Example 1

Induction of neuronal differentiation in neural crest stem cells by forced expression of NGN1.

A retroviral vector harboring an *ngn1* cDNA was constructed so that the NGN1 coding sequence is followed by an internal ribosome entry site (IRES) (from the encephalomyocarditis virus), which in turn is followed by the gene encoding green fluorescent protein (GFP). Transcription of integrated proviral sequences in infected cells thus produces a bi-cistronic mRNA that encodes both NGN1 and GFP. Infected cells can therefore be visualized by virtue of GFP fluorescence (Figure 2, lower, arrows) or by immunostaining with anti-myc tagged GFP or anti-GFP.

Neural crest stem cells (NCSCs), cultured as previously described (Stemple *et al.* 1992. Cell 71:973-985; Lo et al. 1998. Development 125:609-920) were fixed 2.5 days post infection with NGN1-IRES-GFP retrovirus and analyzed at clonal density. The results indicate that virtually all cells that are GFP+ (Fig. 4C, arrows) have a process-bearing, neuronal morphology (Fig. 4A, arrows) and express neurofilament 160 Kd subunit (NF160) (Fig. 4b, arrows). Such differentiation occurs rapidly and is detectable after 2-2.5 days. Cells not expressing GFP have a flat morphology (Figure 4C, Figure 4A) and do not express NF160 (Figure 4B). In high density cultures where neuronal morphology is not easily distinguished (Figure 5A), NGN1-expressing cells (Figure 5C) can be seen to express NeuN (Figure 5B), a neuron-specific nuclear protein. No such induction of neuronal markers is observed when cells are infected with a control virus encoding GFP-IRES-Alkaline Phosphatase.

These data are similar to those obtained with a MASH1-IRES-GFP retrovirus (Fig. 3A-C), except that neuronal differentiation is much more efficient with the NGN1 retrovirus: essentially all NGN1-infected cells express a neuronal phenotype, while only 16% of MASH1-infected cells undergo neurogenesis under these conditions (Lo et al., 1998).

Forced expression of NGN1 thus promotes a core program of neurogenesis in neural crest stem cells. We predict that NGN1 will similarly promote neurogenesis in neural stem cells from the CNS. Thus, introduction of NGN1 coding sequences into undifferentiated neural crest stem cells can be used to efficiently promote neuronal differentiation of these cells, without the need to manipulate their cell culture environment.

Example 2

Expression of neuronal genes in non-neuronal cells by expression of NGN1

Murine NGN1 was expressed in cultured chick embryo fibroblasts (CEFs), using a replication-competent avian retroviral vector (RCAS). In this case, GFP was not used as a marker; rather a myc epitope tag was fused to the *ngn1* coding sequence to allow identification of infected cells by immunocytochemistry using a monoclonal antibody to the myc tag (9E 10). Cells were harvested have 5 days. Culture conditions are cited in Perez et al. (1999) Development 126:1715-1728.

Expression of NGN1 in CEFs caused induction of a number of markers of neuronal differentiation, including neuron-specific beta-tubulin (Figure 6D), neurofilament (NAPA-73; Figure 6B), and NF160 (Figure 6C). In addition the cells displayed morphological changes suggestive of neuronal differentiation. No induction of these markers was detected in control cultures infected with a retrovirus harboring luciferase gene (data not shown).

These data indicate that forced expression of NGN1 can elicit expression of at least some neuronal phenotypic markers even in non-neuronal cells. Thus, introduction of

NGN1 coding sequences into certain non-neuronal cell types, which may be more easily accessed by biopsy than neural stem cells, may be used to promote expression of some neuronal properties which may offer therapeutic benefit in an appropriate transplantation setting. Such an approach would be particularly amenable to autografting.

Example 3

Induction of tyrosine hydroxylase (TH) expression in NCSCs by forced expression of Phox2a

To induce TH expression, cultured rat NCSCs were infected with a retrovirus vector expressing the paired homeodomain protein, Phox2a (Lo *et al.*, Development, 1998. 125:609-620). The Phox2a protein contained a myc epitope tag to permit visualization of the expressed protein in infected cells by immunocytochemistry using an anti-myc monoclonal antibody.

NCSCs were infected and the percentage of retrovirally-infected clones (clones containing any myc-tag positive cells) containing at least one TH+ cell was determined after 96 hours of growth a clonal density under the indicated conditions. Approximately 10% of all infected clones contained detectable levels of TH (Figure 7), and within these clones about 2.5% of all cells were TH-positive (Figure 8). No induction of TH was seen using a control retrovirus encoding a myc-tagged form of GFP (Figure 10F). Inclusion of forskolin in the culture medium (which increases intracellular cAMP) increased the percentage of infected (myc-tag$^+$) clones expressing TH to about 50% (Figure 7), and within these clones almost 15% of the cells were TH$^+$ (Figure 8). In contrast to the result obtained using NGN1, the TH$^+$ cells produced by forced expression of Phox2a did not have a neuronal morphology (Figure 9A) and did not express pan-neuronal markers such as NF160 (not shown). The percentage of infected cells expressing TH could be further increased to almost 35% (Figure 8), by inclusion of additional factors such as GDNF and dexamethasone (DEX) together with forskolin (Lo et al. (1999) Neuron. 22:693-705).

We predict that simultaneous expression of both Phox2a and NGN1 in the same cell would cause the differentiation of neurons that express TH. Such neurons might be useful for transplantation in Parkinson's Disease. The ability to uncouple the expression of neuronal subtype properties from the expression of pan-neuronal properties implies that it should be possible to control the differentiation of neural stem cells to particular neuronal subtypes by expressing in them appropriate combination of transcription factors.

Example 4

Induction of neuronal differentiation in neural crest stem cells by forced expression of NGN1 and Phox2a.

Neural crest stem cells (NCSCs) are co- infected with the NGN1-IRES-GFP and the Phox2a retroviruses described above and grown at clonal density. Virtually all cells that are GFP+ have process-bearing, neuronal morphology, express neurofilament 160 Kd subunit (NF160), NeuN, and TH. The differentiation occurs rapidly and is detectable after 2-2.5 days. The cell is terminally differentiated and have membrane conductance potential similar to catecholamine producing neurons of the PNS.

Example 5

Induction of Brn-3.0 in neural crest cells

Neural crest cells were infected with the retrovirus, NGN1-IRES-GFP, and grown in mass culture. Following 3-5 days of culture, the cells developed a neuronal morphology and expressed Brn-3.0, which is a sensory neuron-specific marker, characteristic of dorsal root ganglion primary sensory neurons. (Greenwood and Anderson. (1999). Development 126: 3543-3559)

What is claimed is:

1. A method of inducing a non-neuronal cell to differentiate into a neuronal cell, comprising:

contacting said non-neuronal cell with an expression vector comprising a neurogenin nucleic acid operatively linked to a promoter functional in said non-neuronal cell, wherein said neurogenin nucleic acid hybridizes under high stringency conditions to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11, or complements thereof, wherein said neurogenin nucleic acid is expressed in said cell.

2. The method according to Claim 1, wherein said non-neuronal cell is an embryonic stem cell.

3. The method according to Claim 1, wherein said non-neuronal cell is a neural stem cell.

4. The method according to Claim 3, wherein said neural stem cell is a neural crest stem cell.

5. The method according to Claim 1, wherein said non-neuronal cell is a fibroblast.

6. The method according to Claim 5, wherein said fibroblast is a chick embryo fibroblast.

7. The method according to Claim 1, wherein said nucleic acid encodes a neurogenin-1.

8. The method according to Claim 1, wherein said expression vector further comprises a sequence encoding a selectable marker.

9. The method according to Claim 8, wherein said selectable marker is a drug resistance marker.

10. The method according to Claim 8, wherein said selectable marker is an epitope tag.

11. The method according to Claim 1, wherein said expression vector comprises a plasmid.

12. The method according to Claim 1, wherein said vector comprises a retrovirus vector.

13. The method according to Claim 1, wherein said vector comprises a herpesvirus vector.

14. A method of inducing a non-neuronal cell to express a core program of neurogenesis, comprising:

contacting said non-neuronal cell with an expression vector comprising a neurogenin nucleic acid operatively linked to a promoter, wherein said neurogenin nucleic acid hybridizes under high stringency conditions to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11, or complements thereof; wherein said neurogenin nucleic acid is expressed in said non-neuronal cell.

15. The method according to Claim 14, wherein said core program of neurogenesis comprises expression of beta-tubulin, neurofilament, NeuN and/or NF160.

16. A method of inducing a non-neuronal cell to express a neuronal marker, comprising:

contacting said non-neuronal cell with an expression vector comprising a

Phox2a nucleic acid operatively linked to a promoter functional in said non-neuronal cell, wherein said Phox2a nucleic acid hybridizes under high stringency conditions to SEQ ID NO:13 or complements thereof, wherein said Phox2a nucleic acid is expressed in said non-neuronal cell.

17. A method of inducing a non-neuronal cell to express a neuronal marker, comprising:

contacting said non-neuronal cell with an expression vector comprising a Phox2b nucleic acid operatively linked to a promoter functional in said non-neuronal cell, wherein said Phox2b nucleic acid hybridizes under high stringency conditions to SEQ ID NO:15 or complements thereof, wherein said Phox2b nucleic acid is expressed in said non-neuronal cell.

18. The method according to Claim 16 or 17, wherein said neuronal marker is an enzyme produced by neurons that synthesize a catecholamine neurotransmitter.

19. A method of inducing a non-neuronal cell to differentiate into a neuron, comprising:

contacting said cell with a neurogenin nucleic acid operatively linked to a first promoter and with a Phox2a nucleic acid operatively linked to a second promoter, wherein said first and second promoters are functional in said non-neuronal cell and said neurogenin nucleic acid hybridizes under high stringency conditions to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11, or complements thereof and said Phox2a nucleic acid hybridizes under high stringency conditions to SEQ ID NO:13 or its complement, wherein said neurogenin and said Phox2a nucleic acids are expressed in said non-neuronal cell.

20. The method according to claim 18, wherein said neuron is a catecholamine-synthesizing neuron.

21. A non-neuronal cell having an neuronal phenotype induced by an

expression vector comprising a neurogenin nucleic acid operatively linked to a first promoter and a Phox2a nucleic acid operatively linked to a second promoter, wherein said first and second promoters are functional in said cell and said neurogenin nucleic acid hybridizes under high stringency conditions to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11, or complements thereof and said Phox2a nucleic acid hybridizes under high stringency conditions to SEQ ID NO:13 or its complement.

22. A method of identifying an agent that modulates neurogenesis in a transformed non-neuronal cell comprising an expression vector comprising a neurogenin nucleic acid operatively linked to a promoter functional in said transformed cell, wherein said neurogenin nucleic acid hybridizes under high stringency conditions to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, or SEQ ID NO:11, or complements thereof comprising the steps of:

a) contacting said transformed cell with said agent; and

b) detecting a modulation in the induction of neurogenesis in said transformed cell.

23. A method of identifying an agent that modulates the induction of a neuronal subtype-specific phenotype in a transformed non-neuronal cell comprising an expression vector comprising a Phox2a nucleic acid operatively linked to a promoter functional in said transformed cell, wherein said Phox2a nucleic acid hybridizes under high stringency conditions to SEQ ID NO:13 or complements thereof comprising the steps of:

a) contacting said transformed cell with said agent; and

b) detecting a modulation in the induction of a neuronal subtype-specific phenotype in said transformed cell.

24. A method of identifying an agent that modulates the induction of a neuronal subtype-specific phenotype in a transformed non-neuronal cell comprising an expression vector comprising a Phox2b nucleic acid operatively linked to a promoter

functional in said transformed cell, wherein said Phox2b nucleic acid hybridizes under high stringency conditions to SEQ ID NO:15 or complements thereof comprising the steps of:

a) contacting said transformed cell with said agent; and

b) detecting a modulation in the induction of a neuronal subtype-specific phenotype in said transformed cell.

(1). Rat Neurogenin cDNA Sequence

```
   1  ATCCGGAGCT GATCTGATCG CCGGCGACAT CAGTCGGGAG ACCAGCCCGG

  51  CGCGTGGCCC CCTGCAGGCG AGGCGAGGAG GCCAAGCCCA TTCCCTCCCT

 101  GAGCCCCTGC GATCTTCCCC GGCCCTCGCG CCTGCAGCAG GCACAGGCTA

 151  GCCCCGGGTC ATACGGACAG TAAGTGCGCT TCGAAGGCCG TGCACTCGGC

 201  CCACATTCAA GCCCTCCAAA CCTCCCGTCC GTCCGTCCGT CCTGCAACGA

 251  TGCCTGCCCC TTTGGAGACC TGTCTCTCTG ACCTCGACTG CGCCAGCAGC

 301  AACAGCGGGA GCGACCTGTC CAGTTTCCTC ACCGACGAGG AGGACTGTGC

 351  CAGGCTCCAG CCCCTAGCTT CCACCTCAGG GCTGTCCGTG CCAGCCCGCA

 401  GGAGCGCGCC CACCCTCTCC GGGGCATCGA ACGTTCCCGG TGGCCAGGAC

 451  GAAGAGCAGG AGCGGCGGCG ACGGCGAGGT CGCGCGCGGG TGCGGTCCGA

 501  GGCGCTGCTG CACTCGCTGC GGAGGAGCCG TCGCGTCAAG GCCAACGATC

 551  GCGAGCGCAA CCGTATGCAT AACCTCAACG CTGCGCTGGA CGCTCTGCGC

 601  AGCGTGCTGC CCTCGTTCCC CGACGACACC AAGCTCACCA AGATTGAGAC

 651  GCTGCGCTTC GCCTACAACT ACATCTGGGC CCTGGCTGAG ACACTGCGCC

 701  TGGCAGATCA AGGGCTCCCG GGGGGCGGTG CCCGGGAGCG CCTCCTGCCT
```

FIG._1A-1

```
751  CCGCAGTGTG TCCCCTGCCT GCCCGGTCCC CCGAGCCCGG CCAGCGATAC
801  AGAGTCCTGG GGCTCCGGGG CCGCTGCCTC CCCCTGCGCT ACTGTGGCGT
851  CACCACTCTC TGACCCCAGT AGTCCCTCGG CTTCAGAAGA CTTCACCTAT
901  GGCCCGGGTG GTCCCCTTTT CTCCTTTCCT GGCCTGCCCA AAGACCTCCT
951  CCATACGACA CCCTGCTTCA TCCCGTACCA CTAGGGCTTT GCAAGACAAC
1001 GTTAATACTT CTTTCCTGCC CCAGTCTATG AGCAATAGAT GGGGGAGCCG
1051 GCTGAAGCCT CGGGGAGCAC CCTTACCCCC AGGTGGATGC TGGGAGCTTT
1101 AAAGAGGGGA GGGATACCTG ACCACTTGCT AGGTTGCCGC ACCCTCGCTG
1151 AGAAGCTGCC CCTCGGACTG TTTCCCCACG CCCCAGCACC GGGCCCCTCC
1201 TGCCCCGCCC CCAGACGGGC TTTCGGTTTT TTTTTTGGAC TTCCTGAACT
1251 TCACAAAACC TCCTTTGTGA CTGGCTCAGA ACTGACCCCA GCCACCACTT
1301 CAGTGTGATT TGGAAAAGGG ACAGATGAGC CCCTGAAGAC GAGGTGAAAA
1351 GTCAATTTTA CAATTTGTAG AACTCTAATG AAGAAAAACG AGCATGAAAA
1401 TTCGGTTTGA GCCGGCTGAC AATACAATGA AAAGGCTTAA AAAAAGGAG
1451 ACACAAGGAG TGGGCTTCAT GCATTATGGA TCCCGACCCC CACCACTGTG
1501 GGCTTGCTCC CGGAAGAACT GAGTGCT
```

FIG._1A-2

```
Met Pro Ala Pro Leu Glu Thr Cys Leu Ser Asp Leu Asp Cys Ala Ser
 1               5                  10                  15

Ser Asn Ser Gly Ser Asp Leu Ser Ser Phe Leu Thr Asp Glu Glu Asp
            20                  25                  30

Cys Ala Arg Leu Gln Pro Leu Ala Ser Thr Ser Gly Leu Ser Val Pro
        35                  40                  45

Ala Arg Arg Ser Ala Pro Thr Leu Ser Gly Ala Ser Asn Val Pro Gly
    50                  55                  60

Gly Gln Asp Glu Glu Gln Glu Arg Arg Arg Arg Arg Gly Arg Ala Arg
65                  70                  75                  80

Val Arg Ser Glu Ala Leu Leu His Ser Leu Arg Arg Ser Arg Arg Val
                85                  90                  95

Lys Ala Asn Asp Arg Glu Arg Asn Arg Met His Asn Leu Asn Ala Ala
            100                 105                 110

Leu Asp Ala Leu Arg Ser Val Leu Pro Ser Phe Pro Asp Asp Thr Lys
        115                 120                 125

Leu Thr Lys Ile Glu Thr Leu Arg Phe Ala Tyr Asn Tyr Ile Trp Ala
    130                 135                 140

Leu Ala Glu Thr Leu Arg Leu Ala Asp Gln Gly Leu Pro Gly Gly Gly
145                 150                 155                 160

Ala Arg Glu Arg Leu Leu Pro Pro Gln Cys Val Pro Cys Leu Pro Gly
                165                 170                 175

Pro Pro Ser Pro Ala Ser Asp Thr Glu Ser Trp Gly Ser Gly Ala Ala
            180                 185                 190

Ala Ser Pro Cys Ala Thr Val Ala Ser Pro Leu Ser Asp Pro Ser Ser
        195                 200                 205

Pro Ser Ala Ser Glu Asp Phe Thr Tyr Gly Pro Gly Gly Pro Leu Phe
    210                 215                 220

Ser Phe Pro Gly Leu Pro Lys Asp Leu Leu His Thr Thr Pro Cys Phe
225                 230                 235                 240

Ile Pro Tyr His
```

FIG._1B

(2). Mouse Neurogenin cDNA Sequence (Open Reading Frame)

1	ATGCCTCCCC CTTTGGAGAC CTGCATCTCT GATCTCGACT
41	GCTCCAGCAG CAACAGCAGC AGCGACCTGT CCAGCTTCCT CACCGACGAG
91	GAGGACTGTG CCAGGCTACA GCCCCTAGCC TCCACCTCGG GGCTGTCCGT
141	GCCAGCCCGG AGGAGCGCTC CCGCCCTCTC CGGGGCATCG AATGTTCCCG
191	GTGCCCAGGA CGAAGAGCAG GAACGGCGCA GGCGGCGAGG TCGCGCTCGG
241	GTGCGGTCCG AGGCTCTGCT GCACTCCCTG CGGAGGAGTC GTCGCGTCAA
291	AGCCAACGAT CGCGAGCGCA ACCGCATGCA CAACCTCAAC GCTGCGCTGG
341	ACGCCTTGCG CAGCGTGCTG CCCTCGTTCC CCGACGACAC CAAGCTCACC
391	AAGATTGAGA CGCTGCGCTT CGCCTACAAC TACATCTGGG CCCTGGCTGA
441	GACACTGCGC CTGGCAGATC AAGGGCTCCC CGGGGGCAGT GCCCGGGAGC
491	GCCTCCTGCC TCCGCAGTGT GTCCCCTGTC TGCCCGGGCC CCCGAGCCCG
541	GCCAGCGACA CTGAGTCCTG GGGTTCCGGG GCCGCTGCCT CCCCCTGCGC
591	CACTGTGGCA TCACCACTCT CTGACCCCAG TAGTCCCTCG GCTTCAGAAG
641	ACTTCACCTA TGGCCCGGGC GATCCCCTTT TCTCCTTTCC TGGCCTGCCC
691	AAAGACCTGC TCCACACGAC GCCCTGTTTC ATCCCATACC ACTAGTAA

FIG._1C

```
Met Pro Pro Pro Leu Glu Thr Cys Ile Ser Asp Leu Asp Cys Ser Ser
 1               5                  10                  15

Ser Asn Ser Ser Ser Asp Leu Ser Ser Phe Leu Thr Asp Glu Glu Asp
            20                  25                  30

Cys Ala Arg Leu Gln Pro Leu Ala Ser Thr Ser Gly Leu Ser Val Pro
        35                  40                  45

Ala Arg Arg Ser Ala Pro Ala Leu Ser Gly Ala Ser Asn Val Pro Gly
    50                  55                  60

Ala Gln Asp Glu Glu Gln Glu Arg Arg Arg Arg Arg Gly Arg Ala Arg
65                  70                  75                  80

Val Arg Ser Glu Ala Leu Leu His Ser Leu Arg Arg Ser Arg Arg Val
                85                  90                  95

Lys Ala Asn Asp Arg Glu Arg Asn Arg Met His Asn Leu Asn Ala Ala
            100                 105                 110

Leu Asp Ala Leu Arg Ser Val Leu Pro Ser Phe Pro Asp Asp Thr Lys
        115                 120                 125

Leu Thr Lys Ile Glu Thr Leu Arg Phe Ala Tyr Asn Tyr Ile Trp Ala
    130                 135                 140

Leu Ala Glu Thr Leu Arg Leu Ala Asp Gln Gly Leu Pro Gly Gly Ser
145                 150                 155                 160

Ala Arg Glu Arg Leu Leu Pro Pro Gln Cys Val Pro Cys Leu Pro Gly
                165                 170                 175

Pro Pro Ser Pro Ala Ser Asp Thr Glu Ser Trp Gly Ser Gly Ala Ala
            180                 185                 190

Ala Ser Pro Cys Ala Thr Val Ala Ser Pro Leu Ser Asp Pro Ser Ser
        195                 200                 205

Pro Ser Ala Ser Glu Asp Phe Thr Tyr Gly Pro Gly Asp Pro Leu Phe
    210                 215                 220

Ser Phe Pro Gly Leu Pro Lys Asp Leu Leu His Thr Thr Pro Cys Phe
225                 230                 235                 240

Ile Pro Tyr His
```

FIG._1D

(3). X-ngnr-1a cDNA Sequence:

1	GCGTGTCACA CGGCAGTTGC ACTCATAATA CACTGTGAGC TGACAGTCGC
51	AACCACGCCC GACAGGGAAC ACGCAGCAAG TCTACTGCAC GACTATAACC
101	CGACGACTCG ACCCAACTCA CCTGCTGCTT CAGGGGCCAA ACACCAAGTT
151	ATAAAGTAAG TAACTTCCAT TGCAACTGCA GCATTGTCAC TTGCGACAGC
201	GCATGAAGTA GTGAGAGGCA CAGACCATGT ACATATATGG GGTTTGTGGT
251	TATTATAGTA AGTGGGATGA TGTTTGGGTT ATTATAGTAA GTGGATGTGA
301	AGTTGTCAGT GCAACATTGG GGCTAACCAT TGGCTGTGTG TTTGCGCTTG
351	TCTAGGATGG TGCTGCTCAA GTGCGAGTAC CGCGATGAAG AGGAGGACCT
401	GACCTCTGCC TCCCCCTGCT CCGTGACCTC CTCTTTCCGT TCCCCGGCGA
451	CGCAGACGTG CAGCTCGGAC GATGAGCAGC TCCTGAGTCC CACCAGCCCG
501	GGACAGCACC AGGGGGAAGA GAACAGCCCG CGATGCAGGA GGAGCCGAGG
551	CCGCGCTCAG GGCAAGAGCG GAGAAACTGT GTTAAAGATC AAGAAGACCC
601	GGCGCGTTAA AGCTAACAAC CGGGAAAGGA ATCGCATGCA CAACCTGAAC
651	TCTGCGCTTG ATTCCCCTCAG GGAAGTGTTG CCCTCTTTAC CTGAAGATGC
701	CAAACTCACC AAGATAGAGA CCTTGCGCTT TGCCTACAAC TACATCTGGG

FIG._1E-1

751 CTCTTAGCGA AACTTTGCGC CTTGGCGACC CAGTGCACCG ATCTGCTTCC

801 ACCCCAGCAG CAGCCATATT GGTGCAGGAC TCCTCTTCAT CCCAGAGCCC

851 CTCCTGGAGC TGCAGCTCGT CCCCTTCTTC CTCTTGTTGC TCCTTCTCCC

901 CGGCCAGCCC TGCCAGCTCC ACCTCGGACA GTATTGAGTC CTGGCAGCCC

951 TCTGAGCTCC ACCTGAACCC CTTCATGTCT GCCAGCAGCG CTTTCATTTG

1001 AACTCCTGTT GGACTATGAT GGATTCTCAC ACTTCCAATT GCTACATATG

1051 AAGAATACCT CAGTGGGGCC CCAGTGCAAA TGATTTTCCT GGGAACCCAG

1101 TTTATTGAGC ATGAGCCCAT ATAGTGTAAT AATATCATCC TGCAGTGACC

1151 AAATTGCACT CTGTGGGTTC TGCTGATGGG GAGAAGTGGG GGGCTTGATC

1201 CCCCTGAGTT TGTGCTTACC TGTATAGCAT TTACTCCCCC TGCTGTCATG

1251 CCCCTGGCAT ATGATGGAGT ACATTGCTGG GTCTATTTTA TTATCAGCAA

1301 TGTGAACTGA AA

FIG._1E-2

```
Met Val Leu Leu Lys Cys Glu Tyr Arg Asp Glu Glu Glu Asp Leu Thr
 1               5                  10                  15

Ser Ala Ser Pro Cys Ser Val Thr Ser Ser Phe Arg Ser Pro Ala Thr
            20                  25                  30

Gln Thr Cys Ser Ser Asp Asp Glu Gln Leu Leu Ser Pro Thr Ser Pro
        35                  40                  45

Gly Gln His Gln Gly Glu Glu Asn Ser Pro Arg Cys Arg Arg Ser Arg
    50                  55                  60

Gly Arg Ala Gln Gly Lys Ser Gly Glu Thr Val Leu Lys Ile Lys Lys
65                  70                  75                  80

Thr Arg Arg Val Lys Ala Asn Asn Arg Glu Arg Asn Arg Met His Asn
                85                  90                  95

Leu Asn Ser Ala Leu Asp Ser Leu Arg Glu Val Leu Pro Ser Leu Pro
            100                 105                 110

Glu Asp Ala Lys Leu Thr Lys Ile Glu Thr Leu Arg Phe Ala Tyr Asn
        115                 120                 125

Tyr Ile Trp Ala Leu Ser Glu Thr Leu Arg Leu Gly Asp Pro Val His
    130                 135                 140

Arg Ser Ala Ser Thr Pro Ala Ala Ala Ile Leu Val Gln Asp Ser Ser
145                 150                 155                 160

Ser Ser Gln Ser Pro Ser Trp Ser Cys Ser Ser Ser Pro Ser Ser Ser
                165                 170                 175

Cys Cys Ser Phe Ser Pro Ala Ser Pro Ala Ser Ser Thr Ser Asp Ser
            180                 185                 190

Ile Glu Ser Trp Gln Pro Ser Glu Leu His Leu Asn Pro Phe Met Ser
        195                 200                 205

Ala Ser Ser Ala Phe Ile
    210
```

FIG._1F

X-ngnr-1b cDNA Sequence:

1	CGAGTGCGCA ACACTTGAGC TGGAGTGCGG GGCGCGTGTC ACACACACAC
51	TGAACTGCCA CTGACACCAG AGACACAGCG AGTGGGAACC CCCTGCTACT
101	ACAGGACTAG GAGAAAAGCC GCACAGCCTG CAGCGCCGCA ACCCGACTCA
151	CCTGCTGCTC CCGGAGCCAC AAGCCTGGCG CACAAGATGG TGCTGCTGAA
201	GTGCGAATAC CGCGATGAGG TGTCGGAACT GACCTCTGTC TCCCCCTGCT
251	CCGTGTCCTC CTCCTCTTCA CACCCGTCCC CGGCGATGCA GACGTGCAGC
301	TCGGACGATG AGCAGCTACA CAGTCCGACA AGCCCGACGC TCACGCACCT
351	GCAGCAGGGA CGGGACCAGG GGGAGGAGAA CAGCCCGCGA TGCAGGAGGA
401	GCCGAGCCCG CGGAGACACC GTGCTGAAGA TCAAGAAGAC CCGGCGCGTT
451	AAAGCCAATA ACCGCGAGAG GAATCGCATG CACCACCTGA ACTATGCGCT
501	CGATTCTCTG AGGGAGGTTC TACCGTCATT ACCCGAAGAC GCCAAACTCA
551	CCAAGATAGA GACCTTGCGC TTTGCCCACA ACTACATCTG GGCTCTTAGC
601	GAAACTTTGC GCCTGGCCGA CCAGCTGCAC GGATCTACTT CCACCCCAGC
651	AGCAGCCATA TTGGTACAGG ACTCCTATCC TTCCCTGAGC CCCTCCTGGA
701	GCTGCAGCTC GTCCCCATCC TCCAACTCTT GCGACTCCTT CTCCCCGACC

FIG._1G-1

751 AGCCCTGCCA GCTCCACCTC GGACAGTATT GAGTACTGGC AGCCCTCTGA

801 GCTCCGCTTG AACCCCTTCA TGTCTGCCCT TTGAACGCAC AGGACTATGG

851 GTGATTTTAA CTTTTTACAC TTTAAATTCC TGCTTCCCAT AAGGGTCAAG

901 TACTGCAGGG GTTACATATC AAGTTTACCT CAGGGGGGGC CACAGCAAAT

951 TCTTTTCCTG GGCCCTAAAA TGTCCTCTGA ATTTGAGCCC ATATAGTGCA

1001 ATGGTATAAC CCTGCAATGG TATAATCCAG CAATGGTATA ATCCTGCATC

1051 GTTACCTAAT TGTACTTTGT GGGGTCTGCT GATGGGGGAC AAGTGTTTGA

1101 CCTGTGTCCA GAGTTTCACA TTTACTCCCC CTTTTGGTAT ATCTCTGGCC

1151 GCAACACTTG CTGTGTCTGT TTCATCGTTA GCTATGTGTA TTAGGAAACT

1201 GTCTATCCCT CATCTGCACC TGTTAGACTA CAGCTACCAA CTTCCTGTTA

1251 CCAGGGGGCT ACTGGGTAAT GTACTTC

FIG._1G-2

MVLLKCEYRDEVSELTSVSPCSVSSSSSHPSPAMQTCSSDDEQL
HSPTSPTLTHLQQGRDQGEENSPRCRRSRARGDTVLKIKKTRRVKANNRERNRMHHLN
YALDSLREVLPSLPEDAKLTKIETLRFAHNYIWALSETLRLADQLHGSTSTPAAAILV
QDSYPSLSPSWSCSSSPSSNSCDSFSPTSPASSTSDSIEYWQPSELRLNPFMSAL

FIG._1H

MDYSYLNSYDSCVAAMEASAYGDFGACSQPGGFQYSPLRPAFPAAGPPCPALGSSNCALGALRDHQP
APYSAVPYKFFPEPSGLHEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARV
QVWFQNRRAKFRKQERAASAKGAAGATGAKKGEARCSSEDDDSKESTCSPTPDSTASLPPPPAPSLA
SPRLSPSPLPAALGSGPGPQPLKGALWAGVAGGGGGGPGTGAAELLKAWQPAEPGPGPFSGVLSSFH
RKPGPALKTNLF

FIG._2B

MYKMEYSYLNSSAYESCMAGMDTSSLASAYADFSSCSQASGFQYNPIRTTFGATSGCPSLTPGSCSLGTLR
DHQSSPYAAVPYKLFTDHGGLNEKRKQRRIRTTFTSAQLKELERVFAETHYPDIYTREELALKIDLTEARV
QVWFQNRRAKFRKQERAAAAAAAAAKNGSSGKKSDSSRDDESKEAKSTDPDSTGGPGPNPNPTPSCGANGG
GGGGPSPAGAPGAAGPGGPGGEPGKGGAAAAAAAAAAAAAAAAAAAGGLAAAGGPGQGWAPGPGPITSIP
DSLGGPFASVLSSLQRPNGAKAALVKSSMF"

FIG._2D

Available Sequence of 30R (Restriction Fragment of Neurogenin2 (ngn2) Genomic Clone;
Includes Full Length Open Reading Frame of ngn2

```
                                                                    CTTAGGAAGCGCCAAGCCCGCGGAGGCGGAGGACACCGTGCTCGGTTCCGGGTGGGGGACA
    1 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 60
                                                                    GAATCCTTCGCGGTTCGGGCGCCTCCTCGCCTCCTGTGGCACGACGAGCCAAGGCCCACCCCCCTGT

                                                                    TTCCCGGACACACACCGGAGCAGCTGCGCCGGAACATTGGAGCCGCGTAGGTAAGTG
   61 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 120
                                                                    AAGGGCCTGTGTGTGGCCTCGTCGTCGACGCGGCCTTGTAACCTCGGCGCATCCATTCAC

                                                                    TGCATGCCGCGCGGCTTTCCATTCGCAGGCAGTGTCCCCACGCAGGCTCACGCCGCCCACGC
  121 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 180
                                                                    ACGTACGGCGCCGAAAGGTAAGCGTCCGTCACAGGGGTGCGTCCGAGTGCGGCGGGTGCG

                                                                    TAACTCCATCGTTTAGACGCAGTGACTTCTGTGACCGGCAGAAGGTGGCTCGAGCCCGGG
  181 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 240
                                                                    ATTGAGGTAGCAAATCTGCGTCACTGAAGACACTGGCCGTCTTCCACCGAGCTCGGGCCC

                                                                    GCGCTCCTCCCCAGCTCTGTCCTCGCCATCTTCGCGAATGCACATTGAGGGAGATGGAGG
  241 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 300
                                                                    CGCGAGGAGGGGTCGAGACAGGAGCGGTAGAAGCGCTTACGTGTAACTCCCTCTACCTCC

                                                                    GGGGGGGGCGGGGCGCGGCCAGCGACACTTTACCCTGTCCATTCTGGGAATAAATTTC
  301 ----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+----------+ 360
                                                                    CCCCCCCCGCCCCGCGCCGGTCGCTGTGAAATGGGACAGGTAAGACCCTTATTTAAAG
```

FIG._1/-1

```
    ATCTGCCTCTTCTTTCTCAGGATGTTCGTCAAATCTGAGACTCTGGAGTTGAAGGAGGAA
361 ---------+---------+---------+---------+---------+---------+ 420
    TAGACGGAGAAGAAAGAGTCCTACAAGCAGTTTAGACTCTGAGACCTCAACTTCCTCCTT

a                          M  F  V  K  S  E  T  L  E  L  K  E  E  -

    GAGGAGGTACTGATGCTGCTGGGCTCGGCTTCCCCGGCCTCGGCGACCCTGACCCCGATG
421 ---------+---------+---------+---------+---------+---------+ 480
    CTCCTCCATGACTACGACGACCCGAGCCGAAGGGGCCGGAGCCGCTGGGACTGGGGCTAC

a     E  E  V  L  M  L  L  G  S  A  S  P  A  S  A  T  L  T  P  M  -

    TCCTCCAGCGCGGACGAGGAGGAGGACGAGGAGCTGCGCCGGCCGGGCTCCGCGCGTGGG
481 ---------+---------+---------+---------+---------+---------+ 540
    AGGAGGTCGCGCCTGCTCCTCCTCCTGCTCCTCGACGCGGCCGGCCCGAGGCGCGCACCC

a     S  S  S  A  D  E  E  E  D  E  E  L  R  R  P  G  S  A  R  G  -

    CAGCGTGGAGCGGAAGCCGAGCAGGGGGTGCAGGGCAGTCCGGCGTCGGGTGCCGGGGGT
541 ---------+---------+---------+---------+---------+---------+ 600
    GTCGCACCTCGCCTTCGGCTCGTCCCCCACGTCCCGTCAGGCCGCAGCCCACGGCCCCCA

a     Q  R  G  A  E  A  E  Q  G  V  Q  G  S  P  A  S  G  A  G  G  -

    TGCCGGCCAGGGCGGCTGCTGGGCCTGATGCACGAGTGCAAGCGTCGCCCGTCGCGCTCA
601 ---------+---------+---------+---------+---------+---------+ 660
    ACGGCCGGTCCCGCCGACGACCCGGACTACGTGCTCACGTTCGCAGCGGGCAGCGCGAGT
```

FIG._1I-2

```
a      C  R  P  G  R  L  L  G  L  M  H  E  C  K  R  R  R  P  S  R  S  -

       CGGGCCGTCTCCCGAGGTGCCAAGAGACGGCGGAGACGGTGCAGCGCATCAAGAAGACCCGC
661    ----------+----------+----------+----------+----------+----------+ 720
       GCCCGGCAGAGGGCTCCACGGTTCTGCCGCCTCTGCCACGTCGCGTAGTTCTTCTGGGCG

a      R  A  V  S  R  G  A  K  K  T  A  E  E  T  V  Q  R  I  K  K  K  T  R  -

       AGGCTCAAGGCCAACAACCGCGAGCGCAACCGCATGCACAACCTAAACGCCGCGCTGGAC
721    ----------+----------+----------+----------+----------+----------+ 780
       TCCGAGTTCCGGTTGTTGGCGCTCGCGTTGGCGTACGTGTTGGATTTGCGGCGCGACCTG

a      R  L  K  A  N  N  R  E  R  N  R  M  H  H  N  L  N  A  A  L  D  -

       GCGCTGCGCGAGGTGCTGCCCACCTTCCCCGAGGATGCCAAGCTCACGAAGATCGAGACG
781    ----------+----------+----------+----------+----------+----------+ 840
       CGCGACGCGCTCCACGACGGGTGGAAGGGGCTCCTACGGTTCGAGTGCTTCTAGCTCTGC

a      A  L  R  E  V  L  P  T  F  P  E  D  A  K  K  L  T  K  I  E  E  T  -

       CTGCGCTTCGCCCACAATTACATCTGGGCGCTCACCGAGACTCTGCGCCTGGCGGACCAC
841    ----------+----------+----------+----------+----------+----------+ 900
       GACGCGAAGCGGGTGTTAATGTAGACCCGCGAGTGGCTCTGAGACGCGGACCGCCTGGTG

a      L  R  F  A  H  N  Y  I  W  A  L  T  E  E  T  L  R  L  A  D  H  -

       TGCGCCGGCGCCGGTGGCCTCCAGGGGGCGCTCTTCACGGAGGCGGTGCTCCTGAGCCCG
901    ----------+----------+----------+----------+----------+----------+ 960
       ACGCGGCCGCGGCCACCGGAGGTCCCCCGCGAGAAGTGCCTCCGCCACGAGGACTCGGGC
```

FIG._1/-3

```
a       C  A  G  A  G  G  G  L  Q  G  A  L  F  T  E  A  V  L  L  S  P  -

          GGAGCTGCGCTCGGCGCCAGCGGGGACAGCCCCTTCTCCACCTTCCTCCTGGAGCTGCACC
      961 ----------+----------+----------+----------+----------+----------+ 1020
          CCTCGACGCGAGCCGCGGTCGCCCCCTGTCGGGAAGAGGTGGAAGGAGGACCTCGACGTGG

a       G  A  A  L  G  A  S  G  D  S  P  S  P  P  S  S  W  S  C  T  -

          AACAGCCCCGGCGTCATCCTCCAACTCCACGTCCCCATACAGCTGCACTTTATCGCCCGCT
     1021 ----------+----------+----------+----------+----------+----------+ 1080
          TTGTCGGGGCCGCAGTAGGAGGTTGAGGTGCAGGGGTATGTCGACGTGAAATAGCGGGCGA

a       N  S  P  A  S  S  S  N  S  T  S  P  Y  S  C  T  L  S  P  A  -

          AGCCCCGGGTCAGACGTGGACTACTGGCAGCCCCCACCTCCGGAGAAGCATCGTTATGCG
     1081 ----------+----------+----------+----------+----------+----------+ 1140
          TCGGGGCCCAGTCTGCACCTGATGACCGTCGGGGGTGGAGGCCTCTTCGTAGCAATACGC

a       S  P  G  S  D  V  D  Y  W  Q  P  P  P  E  K  H  R  Y  A  -

          CCTCACCTGCCCCTCGCCAGGGACTGTATCTAGAGCTGCGGGTCTCCCTCTCTCGTCTCC
     1141 ----------+----------+----------+----------+----------+----------+ 1200
```

FIG._11-4

```
                                                GGAGTGGACGGGGAGCGGTCCCTGACATAGATCTCGACGCCCAGAGGGAGAGAGCAGAGG

a                                       P  H  L  P  L  A  R  D  C  I  *

                                          TACCCGGGCCCTCCTTCTCCCATCCTTCTCCCGCCCCCCACCCTCCACGCCCCGGAATCCAC
                                 1201 ----------+----------+----------+----------+----------+----------+ 1260
                                          ATGGGCCCGGGAGGAAGAGGGTAGAAGAGGGCGGGGGGTGGGAGGTGCGGGGCCTTAGGTG

                                          TTCACAGAACAGAAGTTGGCCCTTTGCAATCCCCTCCGCGGCTGGTGCTTCGGGGGTTGG
                                 1261 ----------+----------+----------+----------+----------+----------+ 1320
                                          AAGTGTCTTGTCTTCAACCGGGAAACGTTAGGGGAGGCGCCGACCACGAAGCCCCCCAACC

                                          AAAACAACTCTGGTTTATTGAAATTAAGATTTTGGTCAAAAAGAATATGCTTTTTGGAAT
                                 1321 ----------+----------+----------+----------+----------+----------+ 1380
                                          TTTTGTTGAGACCAAATAACTTTAATTCTAAAACCAGTTTTTCTTATACGAAAAACCTTA

                                          TGGGG
                                 1381 ----- 1385
                                          ACCCC
```

FIG._1/5

Available Sequence of 17RB And 17B (Restriction Fragments of Neurogenin3 (ngn3) Genomic Clone; Includes Full Length Open Reading Frame of ngn3

```
       ATTCTTTTGAGTCGGGAGAACTAGGTAACAATTCGGAAACTCCAAAGGGTGGATGAGGGG
     1 ---------+---------+---------+---------+---------+---------+ 60
       TAAGAAAACTCAGCCCTCTTGATCCATTGTTAAGCCTTTGAGGTTTCCCACCTACTCCCC

       CGCGCGGGGTGTGTGTGGGGGATACTCTGGTCCCCCGTGCAGTGACCTCTAAGTCAGAGG
    61 ---------+---------+---------+---------+---------+---------+ 120
       GCGCGCCCCACACACACCCCCTATGAGACCAGGGGGCACGTCACTGGAGATTCAGTCTCC

       CTGGCACACACACACCTTCCATTTTTTCCCAACCGCAGGATGGCGCCTCATCCCTTGGAT
   121 ---------+---------+---------+---------+---------+---------+ 180
       GACCGTGTGTGTGTGGAAGGTAAAAAAGGGTTGGCGTCCTACCGCGGAGTAGGGAACCTA

                                                 M   A   P   H   P   L   D   -

       GCGCTCACCATCCAAGTGTCCCCAGAGACACAACAACCTTTTCCCGGAGCCTCGGACCAC
   181 ---------+---------+---------+---------+---------+---------+ 240
       CGCGAGTGGTAGGTTCACAGGGGTCTCTGTGTTGTTGGAAAAGGGCCTCGGAGCCTGGTG

a          A   L   T   I   Q   V   S   P   E   T   Q   Q   P   F   P   G   A   S   D   H   -

       GAAGTGCTCAGTTCCAATTCCACCCCACCTAGCCCCACTCTCATACCTAGGGACTGCTCC
   241 ---------+---------+---------+---------+---------+---------+ 300
       CTTCACGAGTCAAGGTTAAGGTGGGGTGGATCGGGGTGAGAGTATGGATCCCTGACGAGG

a          E   V   L   S   S   N   S   T   P   P   S   P   T   L   I   P   R   D   C   S   -
```

FIG._1J-1

```
    GAAGCAGAAGTGGGTGACTGCCGAGGGACCTCGAGGAAGCTCCGCGCCCGACGCGGAGGG
301 ---------+---------+---------+---------+---------+---------+ 360
    CTTCGTCTTCACCCACTGACGGCTCCCTGGAGCTCCTTCGAGGCGCGGGCTGCGCCTCCC

a     E  A  E  V  G  D  C  R  G  T  S  R  K  L  R  A  R  R  G  G  -

    CGCAACAGGCCCAAGAGCGAGTTGGCACTCAGCAAACAGCGAAGAAGCCGGCGCAAGAAG
361 ---------+---------+---------+---------+---------+---------+ 420
    GCGTTGTCCGGGTTCTCGCTCAACCGTGAGTCGTTTGTCGCTTCTTCGGCCGCGTTCTTC

a     R  N  R  P  K  S  E  L  A  L  S  K  Q  R  R  S  R  R  K  K  -

    GCCAATGATCGGGAGCGCAATCGCATGCACAACCTCAACTCGGCGCTGGATGCGCTGCGC
421 ---------+---------+---------+---------+---------+---------+ 480
    CGGTTACTAGCCCTCGCGTTAGCGTACGTGTTGGAGTTGAGCCGCGACCTACGCGACGCG

a     A  N  D  R  E  R  N  R  M  H  N  L  N  S  A  L  D  A  L  R  -

    GGTGTCCTGCCCACCTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCCTGCGCTTC
481 ---------+---------+---------+---------+---------+---------+ 540
    CCACAGGACGGGTGGAAGGGCCTACTGCGGTTTGAATGTTTCTAGCTCTGGGACGCGAAG
```

FIG._1J-2

```
a      G  V  L  P  T  F  P  D  D  A  K  L  T  K  I  E  T  L  R  F  -

       GCCCACAACTACATCTGGGCACTGACTCAGACGCTGCGCATAGCGGACCACAGCTTCTAT
541    ---------+---------+---------+---------+---------+---------+ 600
       CGGGTGTTGATGTAGACCCGTGACTGAGTCTGCGACGCGTATCGCCTGGTGTCGAAGATA

a      A  H  N  Y  I  W  A  L  T  Q  T  L  R  I  A  D  H  S  F  Y  -

       GGCCCGGAGCCCCCTGTGCCCTGTGGAGAGCTGGGGAGCCCCGGAGGTGGCTCCAACGGG
601    ---------+---------+---------+---------+---------+---------+ 660
       CCGGGCCTCGGGGGACACGGGACACCTCTCGACCCCTCGGGGCCTCCACCGAGGTTGCCC

a      G  P  E  P  P  V  P  C  G  E  L  G  S  P  G  G  G  S  N  G  -

       GACTGGGGCTCTATCTACTCCCCAGTCTCCCAAGCGGGTAACCTGAGCCCCACGGCCTCA
661    ---------+---------+---------+---------+---------+---------+ 720
       CTGACCCCGAGATAGATGAGGGGTCAGAGGGTTCGCCCATTGGACTCGGGGTGCCGGAGT

a      D  W  G  S  I  Y  S  P  V  S  Q  A  G  N  L  S  P  T  A  S  -

       TTGGAGGAATTCCCTGGCCTGCAGGTGCCCAGCTCCCCATCCTATCTGCTCCCGGGAGCA
721    ---------+---------+---------+---------+---------+---------+ 780
       AACCTCCTTAAGGGACCGGACGTCCACGGGTCGAGGGGTAGGATAGACGAGGGCCCTCGT

a      L  E  E  F  P  G  L  Q  V  P  S  S  P  S  Y  L  L  P  G  A  -

       CTGGTGTTCTCAGACTTCTTGTGA
781    ---------+---------+---- 804
       GACCACAAGAGTCTGAAGAACACT

a      L  V  F  S  D  F  L  *  -
```

FIG._1J-3

```
   1 agctcgaatt ccgttgctgt cgggccgatg gactactcct acctcaattc gtacgattcg
  61 tgcgtggcgg ccatggaggc gtccgcctac ggtgacttcg gcgcctgcag ccagcctgga
 121 ggcttccaat acagtcccct gcggcctgcc ttccccgccg ctgggccacc ttgccccgcg
 181 ctcggctcct ccaactgtgc gcttggcgcc ctacgcgacc accagcccgc accctactcg
 241 gcagttcctt acaagttctt cccggagccg tccggccctgc atgagaagcg caagcagcgg
 301 cgcatccgca caacgttcac gagtgctcag ctcaaggagt tggagcgcgt cttcgccgag
 361 acccactacc ccgacattta cactcgcgag gaactggcgc tcaagatcga cctcactgag
 421 gctcgcgtgc aggtctggtt ccagaaccgc cgggccaagt tccgcaaaca ggagcgcgcg
 481 gccagcgcca aaggcgcggc gggagcgacg ggcgccaaaa agggcgaggc gcgttgctcg
 541 tcggaggacg acgactccaa ggagtccacg tgcagcccca cgcccgacag caccgcgtcg
 601 ctgccgccgc cgcctgcacc cagcctggcc agcccgcgtc tgagccccag ccctctgccc
 661 gccgcgctgg gctccgggcc cgggccccag ccgctcaagg gagcgttgtg ggcagggggtg
 721 gcgggcggtg gaggtggcgg ccccggcacg ggcgcagcgg agctgcttaa ggcctggcag
 781 ccggcggaac ccgggccagg tcccttctct ggagttctgt cctcctttca ccggaagccc
 841 ggcccccgccc tgaagacaaa cctcttctag ccgcgggcgt ctgtaggcaa ccagcctgcc
 901 ccgagagaga caccctcccc cttctggacc tggcattatc cctccctatc ccggcagcct
 961 gcctggaaac tccccgtcgt ccccactacc cagtgtctga tccctagacc tggccccccct
1021 tcgtggtaaa acaagccagc gccactctgg tctggagtac taatcaccgt gccgcccctt
1081 cagggcggcc ggaagcccctt tcttgctagg ctttctttagg aacagggatc aaattacacc
1141 tgtccctcac tcagtgccca atcataaagg gtccctaagaa gccgagccaa cagctcctag
1201 actttcagc tagctgggcc actcattcct tgaaatcaag caacctgaag agtcccaccg
1261 ccaatcccac ccttaacgag tcacctccca tccctagcca gtatggcgca gaggttagac
1321 actagagggg aagagccgtc ggggaacgga acaaaatggt ttccttttc ctttatttt
1381 tctttgaaaa acgtgtaatt tattaaggtg atttgctcaa tccaaataaa acttaattta
1441 ttgaagacaa aaaaaaaaa aaa
```

FIG._2A

```
1    aagcagagcc aagtttatta tgaggactat atactctaga gacctcagac aaggcatctc
61   acaggaggct tttcataaa actaggctcg gctggtagta aggaggccag tgtggaggca
121  ggcgttgagc agtgcacatc tccccactcc agccaccgtc tccacatcca tcttttattt
181  catttttcca cttggctgag ccatccagaa cctttttcaat gtataaaatg gaatattctt
241  acctcaattc ctctgcctac gagtcctgta tggccgggat ggatacctcc agcttggctt
301  cagcatatgc agacttcagt tcctgcagcc aggccagtgg cttccagtat aacccgataa
361  ggaccacttt tgggccacg tccggttgcc cgtccctcac gccgggatcc tgcagccctgg
421  gcaccctcag ggaccaccag agcagtccgt acgccgcagt tccatacaaa ctcttcaccg
481  accacggcgg cctcaacgag aaacgcaagc agcggcgcat ccgcaccacc ttcacaagcg
541  cgcagctcaa agagttggag agggtctttcg ctgagacgca ctaccctgac atctacacca
601  gggaagagct agcactgaag atcgacctca ccgaggcgag agtccaggtg tggttccaga
661  accgccgggc taagtttcgc aagcaggagc gcgcagccgc cgctgctgcc gccgccgcca
721  aaaacggctc ctccgggaag aagtctgact cctcccggga cgacgagagc aaagaggcca
781  agagcactga tcccgacagc actgggggcc cgggacccaa ccccaacccg accccagct
841  gtgggcaaa tggcggcggt ggcggagggc ccagcccagc cggagctccg gggcgggcgg
901  gcccgggggg ccgggaggc gaacccggca agggcggtgc ggctgcgcgct gctgctgcag
961  ctgcagcggc tgcagcggcg gctgcggctg cggcggcggg aggcctggct gcggccggag
021  gccccggaca aggctgggct cctggccccgg gccccatcac ctccatccca gattctcttg
081  gggggccctt cgccagcgtc ttatctttcgc tccaaagacc caacggtgcc aaagccgcct
141  tagtgaagag cagtatgttc tgatctgcga tctgcggcgg cggcgggcga gcccgggggcc
201  cgggctggcg agtgggcgag tgggtagacc caaggctagt gtcgctgctg tcagtggctt
261  ttttttttt ttcatcgaag gcctaaaatg atcgcgattc gaaaaaacaa cgagaaaatg
321  acgtcgctcc atttcaaccc cactcctacc cctttcctca ccccaaaca aagcaaaaca
381  aaacaaacaa aaatcttcac ctgccttggg ctgcgcatag gacagggctc cacctgctgc
441  ccaagggatg tgagcttgga cttcgggggca ctctcagggg gctgtgtctg agtgacgggt
501  gtatgtctgt ctcagagaat gtgtgtctgt ggccccaagt aggtgatagg gagagacggg
561  ggcaccacaa ccaactcagt gacttcttag gaaaaaaaaa aaaa
```

FIG._2C

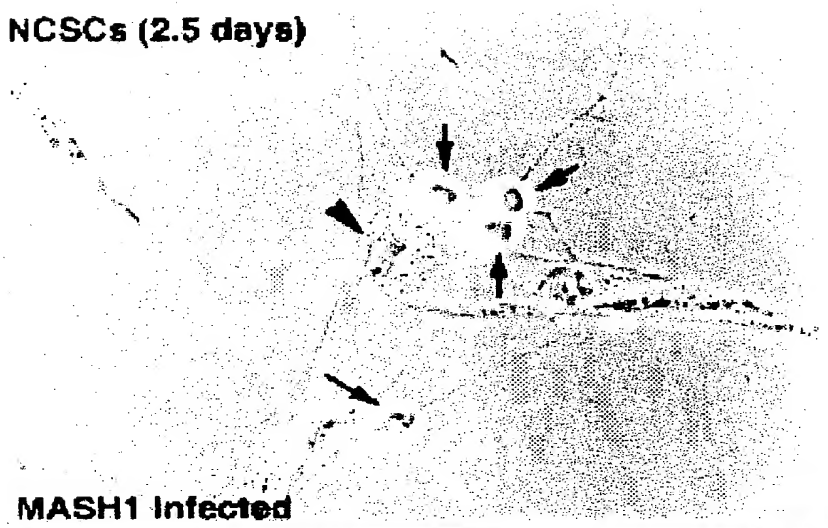

FIG._3A

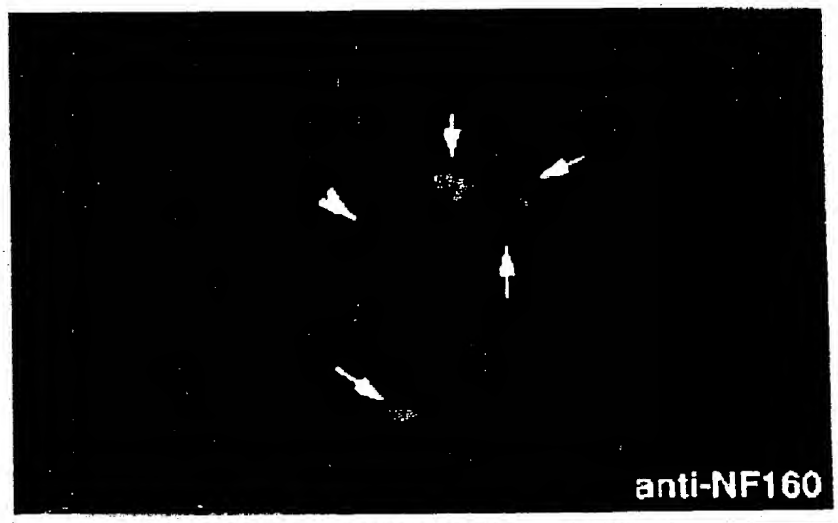

FIG._3B

FIG._3C

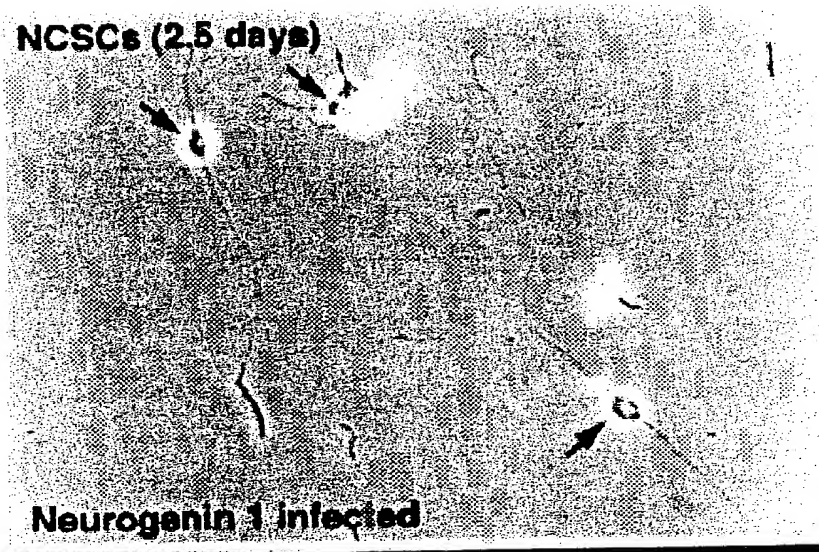

FIG._4A

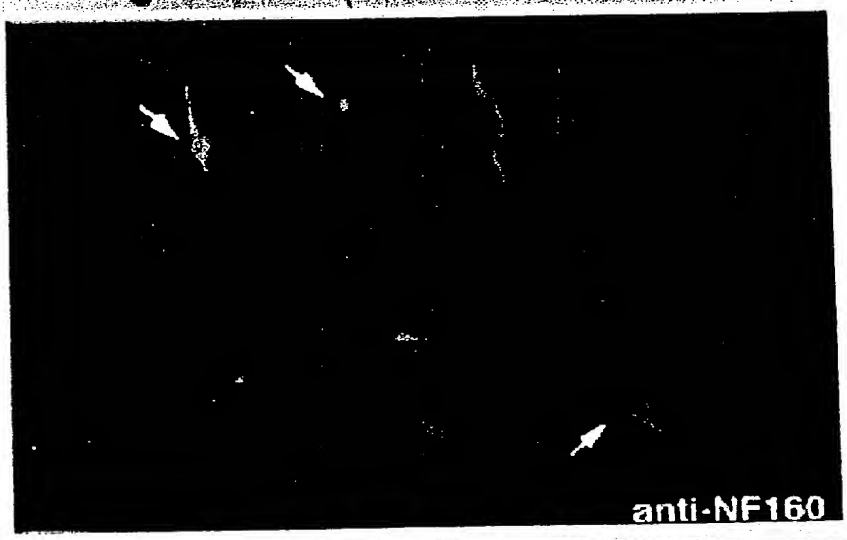

FIG._4B

FIG._4C

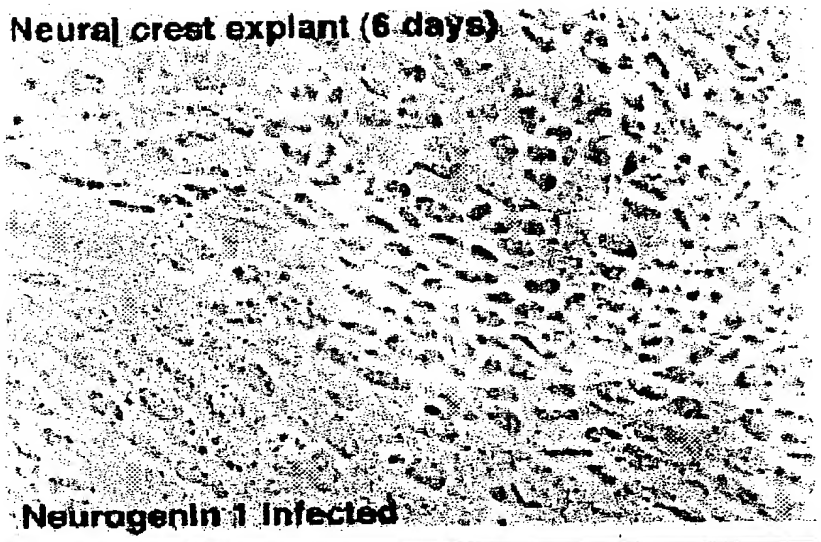

FIG._5A

FIG._5B

FIG._5C

FIG._6A

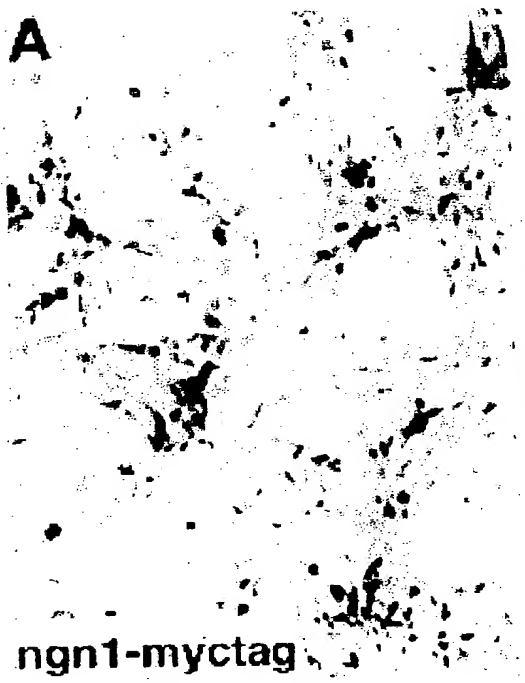

FIG._6B

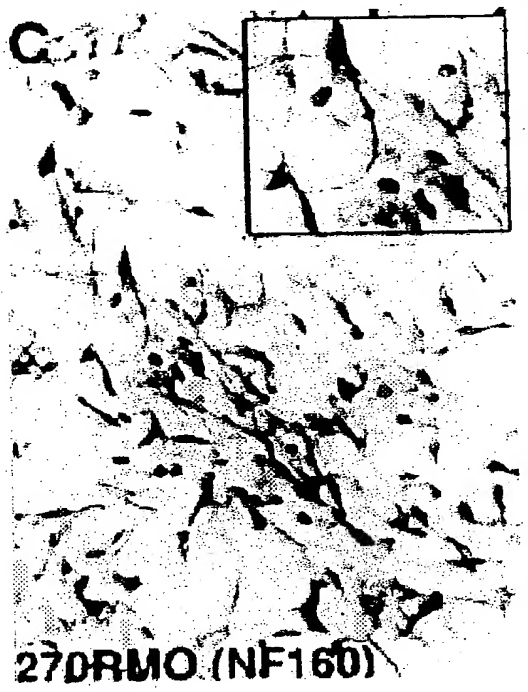

FIG._6C

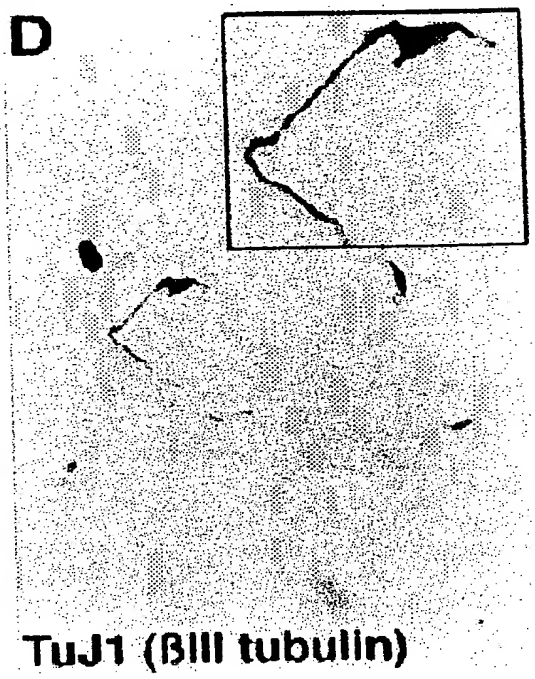

FIG._6D

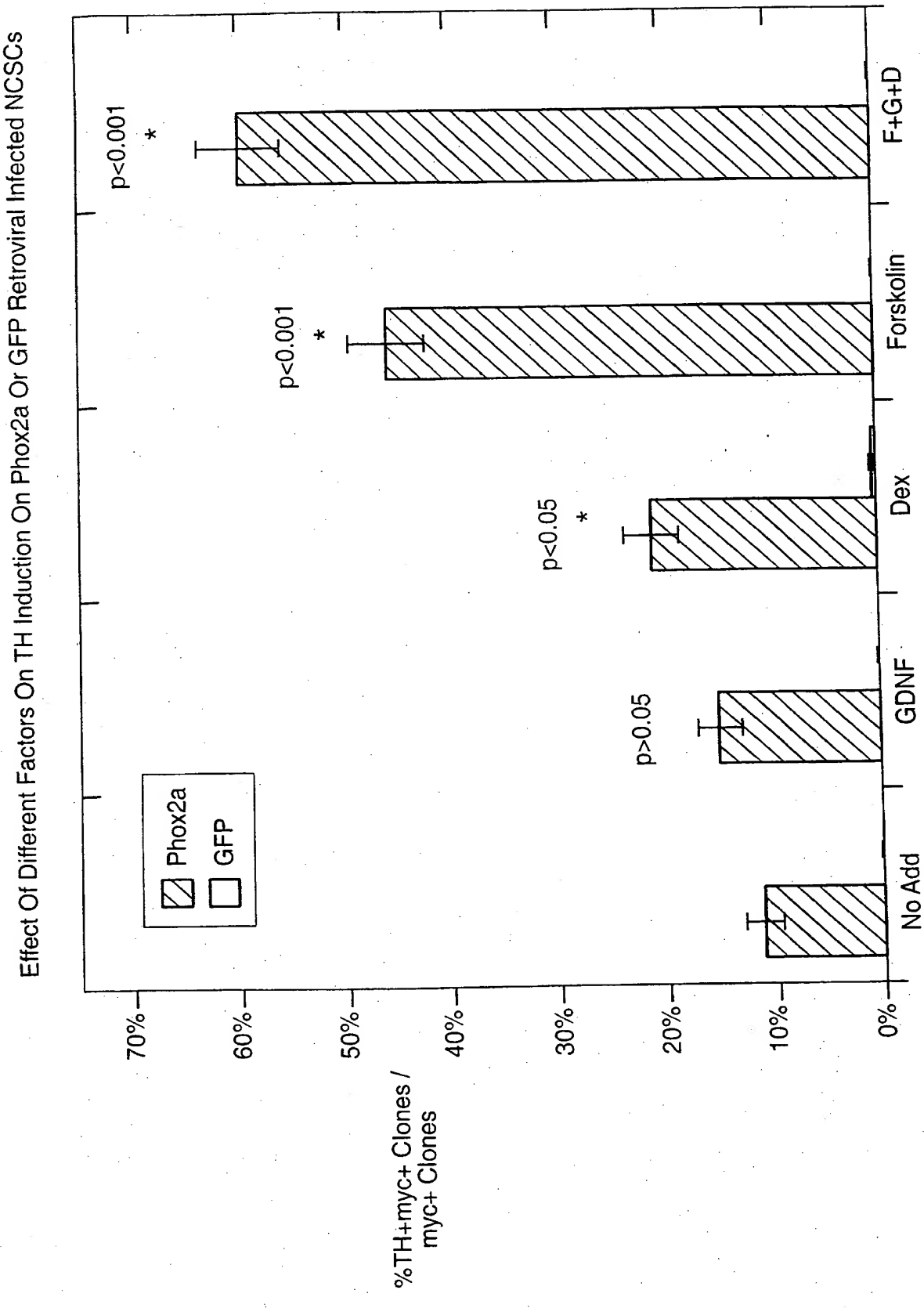

FIG._7

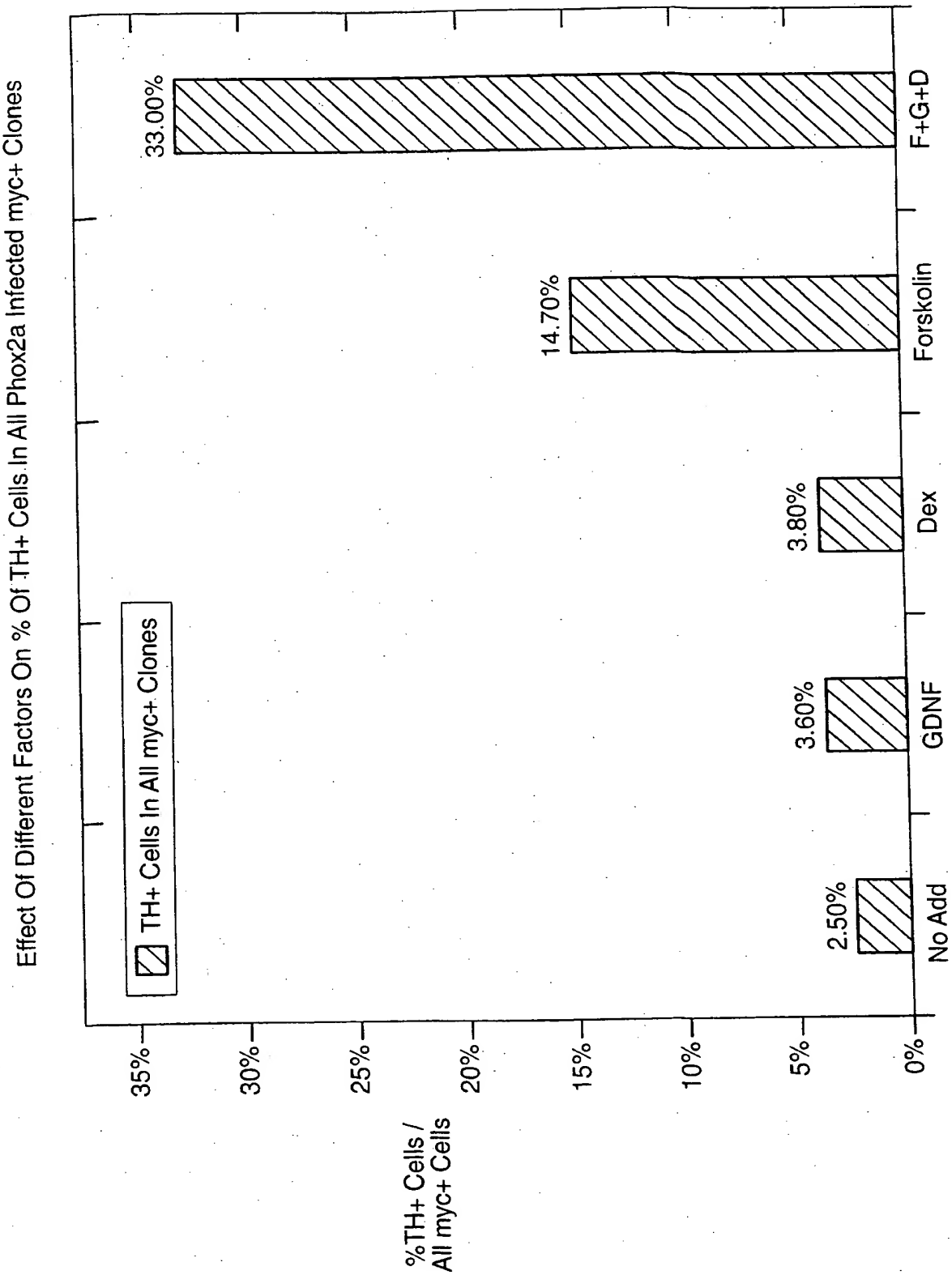

FIG._8

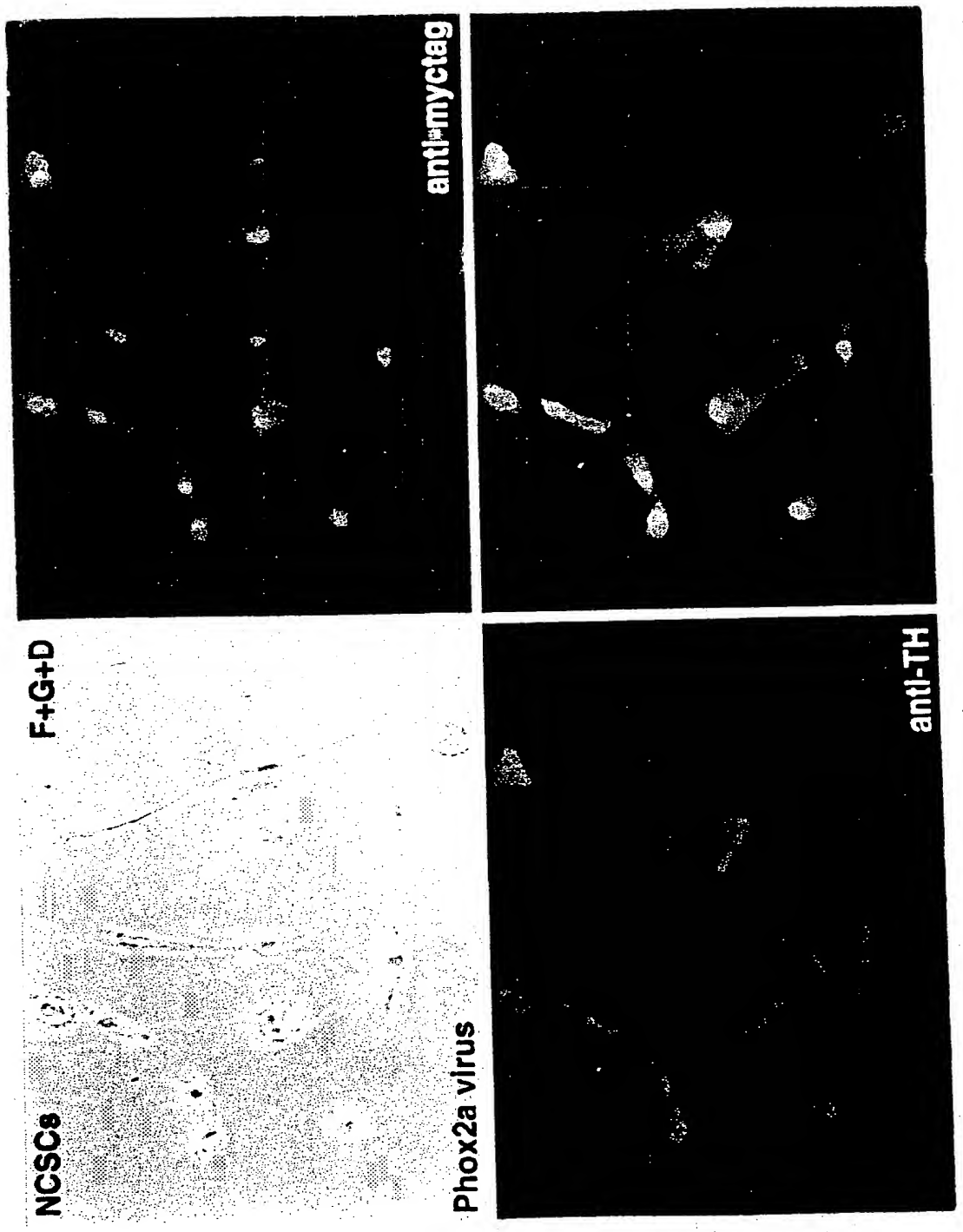

FIG._9A

FIG._9B

FIG._9C

FIG._9D

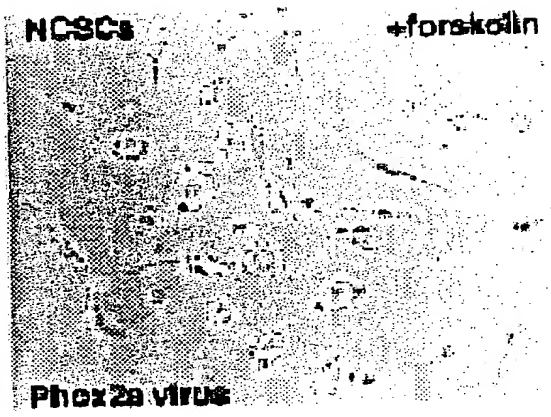

FIG._10A

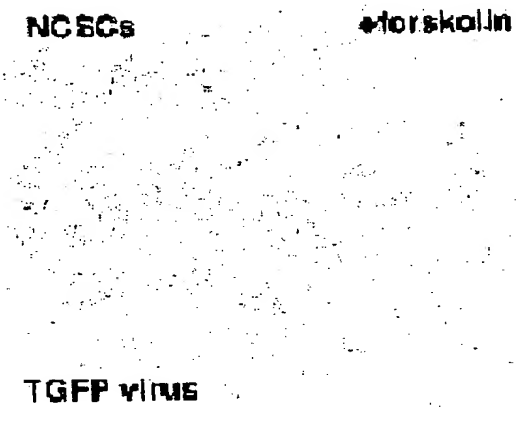

FIG._10B

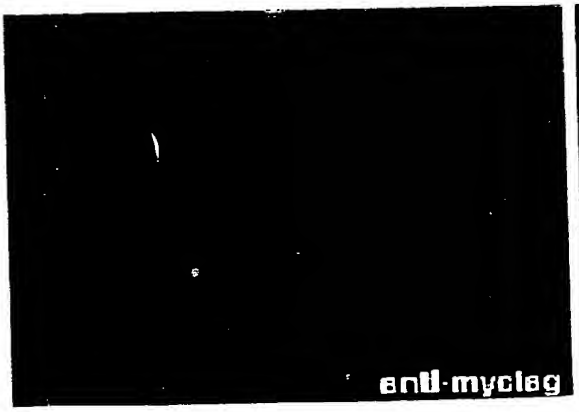

FIG._10C

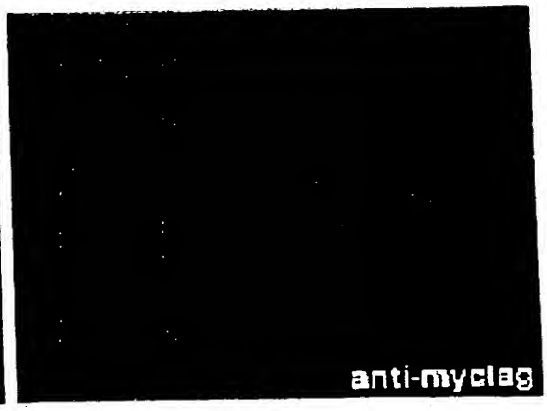

FIG._10D

FIG._10E

FIG._10F

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
C12N 5/06, 15/85, C07K 14/475
C12N 15/66, 15/86, G01N 33/53

A3

(11) International Publication Number: **WO 00/09676**

(43) International Publication Date: 24 February 2000 (24.02.00)

(21) International Application Number: PCT/US99/18525

(22) International Filing Date: 13 August 1999 (13.08.99)

(30) Priority Data:
60/096,630 14 August 1998 (14.08.98) US

(71) Applicant: CALIFORNIA INSTITUTE OF TECHNOLOGY [US/US]; 1200 East California Boulevard, Pasadena, CA 91125 (US).

(72) Inventors: ANDERSON, David, J.; 2891 Mount Curve Avenue, Pasadena, CA 91001 (US). LO, Li-Ching; 525 West Las Tunas Drive, Arcadia, CA 91006 (US).

(74) Agents: TRECARTIN, Richard, F. et al.; Flehr Hohbach Test Albritton & Herbert LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).

(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(88) Date of publication of the international search report:
8 June 2000 (08.06.00)

(54) Title: METHODS OF FORMING NEURONS

(57) Abstract

The invention relates to novel methods of inducing non-neuronal cells to differentiate into neurons and to methods of inducing non-neuronal cells to express a neuronal subtype-specific marker.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

In ıtional Application No
PCT/US 99/18525

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N5/06 C12N15/85 C07K14/475 C12N15/66 C12N15/86
G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 98 13491 A (CALIFORNIA INST OF TECHN) 2 April 1998 (1998-04-02) page 4, line 12 -page 7, line 2; examples 1-4 --- -/--	1-4,7,8, 14,22 10,12, 15-17, 23,24

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 February 2000	09/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Mateo Rosell, A.M.

1

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	LO L ET AL., : "MASH1 activates expression of the paired homeodomain transcription factor Phox2a, and couples pan-neuronal and subtype-specific components of autonomic neuronal identity" DEVELOPMENT, vol. 125, 22 January 1998 (1998-01-22), pages 609-620, XP002130309 cited in the application abstract page 610, right-hand column -page 611, left-hand column page 614, right-hand column -page 619, left-hand column	10,12, 15,16,23
Y	VALARCHE I ET AL., : "The mouse homeodomain protein Phox2 regulates Ncam promoter activity in concert with Cux/CDP and is a putative determinant of neurotransmitter phenotype" DEVELOPMENT, vol. 119, 1993, pages 881-896, XP002130310 cited in the application the whole document -& DATABASE EMBL NUCLEOTIDE AND PROTEIN SEQUENCES,20 September 1993 (1993-09-20), XP002130311 HINXTON, GB AC = X75014. M.musculus Phox2 mRNA for homeodomain protein abstract	16,23
Y	PATTYN A ET AL., : "Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis" DEVELOPMENT, vol. 124, 1997, pages 4065-4075, XP002130312 cited in the application the whole document Fig 1	17,24
X	MA Q ET AL: "IDENTIFICATION OF NEUROGENIN, A VERTEBRAE NEURONAL DETERMINATION GENE" CELL,US,CELL PRESS, CAMBRIDGE, NA, vol. 87, no. 1, 4 October 1996 (1996-10-04), page 43-52 XP002071056 ISSN: 0092-8674 cited in the application the whole document Fig 1	1,2,8, 11,22

-/--

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	MCCORMICK M B ET AL: "NEUROD2 AND NEUROD3: DISTINCT EXPRESSION PATTERNS AND TRANSCRIPTIONAL ACTIVATION POTENTIALS WITHIN THE NEUROD GENE FAMILY" MOLECULAR AND CELLULAR BIOLOGY,US,WASHINGTON, DC, vol. 16, no. 10, October 1996 (1996-10), page 5792-5800 XP002071058 ISSN: 0270-7306 cited in the application the whole document Fig 2 ---	1,2,8, 10,11,22
X	SOMMER L ET AL: "NEUROGENINS, A NOVEL FAMILY OF ATONAL-RELATED BHLH TRANSCRIPTION FACTORS, ARE PUTATIVE MAMMALIAN NEURONAL DETERMINATION GENES THAT REVEAL PROGENITOR CELL HETEROGENEITY IN THE DEVELOPING CNS AND PNS" MOLECULAR AND CELLULAR NEUROSCIENCES,US,SAN DIEGO, vol. 8, no. 4, 1996, page 221-241 XP002072615 ISSN: 1044-7431 cited in the application the whole document Fig 1 ---	1
A	SHIMIZU C ET AL: "MATH-2, A MAMMALIAN HELIX-LOOP-HELIX FACTOR STRUCTURALLY RELATED TO THE PRODUCT OF DROSOPHILA PRONEURAL GENE ATONAL, IS SPECIFICALLY EXPRESSED IN THE NERVOUS SYSTEM" EUROPEAN JOURNAL OF BIOCHEMISTRY,DE,BERLIN, vol. 229, no. 1, 1 April 1995 (1995-04-01), page 239-248 XP002039567 ISSN: 0014-2956 the whole document ---	1
A	SOMMER L ET AL: "THE CELLULAR FUNCTION OF MASH1 IN AUTONOMIC NEUROGENESIS" NEURON,US,CAMBRIDGE, MA, vol. 15, no. 6, December 1995 (1995-12), page 1245-1258 XP002071060 cited in the application the whole document --- -/--	1

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	TURNER D L ET AL: "EXPRESSION OF ACHAETE-SCUTE HOMOLOG 3 IN XENOPUS EMBRYOS CONVERTS ECTODERMAL CELLS TO A NEURAL FATE" GENES AND DEVELOPMENT,US,COLD SPRING HARBOR, NY, vol. 8, page 1434-1447 XP002071061 ISSN: 0890-9369 the whole document	1
P,X	PATTYN A ET AL.: "The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives" NATURE, vol. 399 (6734), 27 May 1999 (1999-05-27), page 366-370 XP000876704 the whole document	17,24
P,X	PEREZ SE ET AL., : "Early specification of sensory neuron fate revealed by expression and function of neurogenins in the chick embryo" DEVELOPMENT, vol. 126, no. 8, 17 March 1999 (1999-03-17), pages 1715-1728, XP000879058 cited in the application page 1716 -page 1717 whole document	1-12,14

INTERNATIONAL SEARCH REPORT

Information on patent family members

Ir ational Application No
PCT/US 99/18525

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9813491 A	02-04-1998	AU 4497297 A	17-04-1998
		EP 0946720 A	06-10-1999